DOCUMENT RESUME

ED 060 686	FL 002 889

AUTHOR	Bond, Zinny S.; And Others
TITLE	Working Papers in Linguistics No. 9.
INSTITUTION	Ohio State Univ., Columbus. Computer and Information Science Research Center.
SPONS AGENCY	National Science Foundation, Washington, D.C. Office of Science Information Services.
REPORT NO	OSU-CISRC-TR-71-8
PUB DATE	Jul 71
NOTE	246p.
EDRS PRICE	MF-$0.65 HC-$9.87
DESCRIPTORS	Artificial Speech; *Auditory Perception; *Comparative Analysis; Dialects; English; Estonian; Experiments; Hungarian; Language; *Language Research; Latvian; *Linguistic Theory; Lithuanian; Morphology (Languages); Mutual Intelligibility; Phonemics; Phonology; *Speech; Suprasegmentals; Syntax

ABSTRACT

This volume of working papers includes eight papers dealing mainly with experimental topics in linguistics and in phonology and speech. The following articles are in the collection: "Units in Speech Perception," "The Temporal Realization of Morphological and Syntactic Boundaries," "Comparison of Controlled and Uncontrolled Normal Speech Rate," "Word Unit Temporal Compensation," "Relative Intelligibility of Five Dialects of English," "Intensity and Duration Analysis of Hungarian Secondary Stress," "Experiments with Synthetic Speech Concerning Quantity in Estonian," "Phonological Rules in Lithuanian and Latvian." Titles of working papers appearing in previous collections are provided. (VM)

ED 060686

(OSU-CISRC-TR-71-8)

WORKING PAPERS IN LINGUISTICS NO. 9

by

Zinny S. Bond, Richard Gregorski,
Andrew Kerek, Ilse Lehiste,
Linda Shockey, and Mary V. Wendell

Work performed under

Grant No. 534.1, National Science Foundation

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Computer and Information Science Research Center
The Ohio State University
Columbus, Ohio 43210
July 1971

FL 002 889

Foreword

The Computer and Information Science Research Center of The Ohio State University is an inter-disciplinary research organization which consists of the staff, graduate students, and faculty of many University departments and laboratories. This report presents research accomplished in cooperation with the Department of Linguistics.

The work of the Center is largely supported by government contracts and grants. The preparation of four of the papers contained in this report was partly supported by the Office of Science Information Service, National Science Foundation under Grant No. GN-534.1.

Ilse Lehiste
CISRC Sub-Project Supervisor

Marshall Yovits
Director
CISRC

Table of Contents

Foreword ii

List of Working Papers in Linguistics iv

Introduction vii

Zinny S. Bond, Units in Speech Perception viii

Ilse Lehiste, "The Temporal Realization of Morphological and Syntactic Boundaries" 113

Richard Gregorski, Linda Shockey, and Ilse Lehiste, "Comparison of Controlled and Uncontrolled Normal Speech Rate" 131

Linda Shockey, Richard Gregorski, and Ilse Lehiste, "Word Unit Temporal Compensation" 145

Mary Virginia Wendell, "Relative Intelligibility of Five Dialects of English" 166

Richard Gregorski and Andrew Kerek, "Intensity and Duration Analysis of Hungarian Secondary Stress" 192

Ilse Lehiste, "Experiments with Synthetic Speech Concerning Quantity in Estonian" 199

Zinny S. Bond, "Phonological Rules in Lithuanian and Latvian" 218

List of WORKING PAPERS IN LINGUISTICS

*No. 1 (December, 1967)

"The Grammar of 'Hitting' and 'Breaking'," Charles J. Fillmore, pp. 9-29. (To appear in Studies in English Transformational Grammar, R. Jacobs and P. Rosenbaum, eds., Ginn-Blaisdell.)

"The English Preposition WITH", Gregory Lee, pp. 30-79.

"Relative Clauses and Conjunctions", Sandra Annear Thompson, pp. 80-99.

"On Selection, Projection, Meaning, and Semantic Content", D. Terence Langendoen, pp. 100-109.

"Some Problems of Derivational Morphology", Sandra Annear Thompson and Dale Elliott, pp. 110-115.

"The Accessibility of Deep (Semantic) Structures", D. Terence Langendoen, pp. 118-127. (To appear in Studies in English Transformational Grammar, R. Jacobs and P. Rosenbaum, eds., Ginn-Blaindell.)

"Review of Haim Gaifman, 'Dependency Systems and Phrase-Structure Systems', Information and Control 8 (1965), pp. 304-337", James T. Heringer, pp. 128-136.

"Diphthongs Versus Vowel Sequences in Estonian", Ilse Lehiste, pp. 138-148. (Also in Proceedings of the VI International Congress of Phonetic Sciences, Prague (1967), Academia: Prague, 1970, pp. 539-544.)

*No. 2 (November, 1968) (OSU-CISRC-TR-68-3)

"Lexical Entries for Verbs", Charles J. Fillmore, pp. 1-29. (Also in Foundations of Language 4 (1968), pp. 373-393.)

"Review of Componential Analysis of General Vocabulary: The Semantic Structure of a Set of Verbs in English, Hindi, and Japanese, Part II, by Edward Herman Bendix. I.J.A.L. Vol. 32, No. 2, Publication 41, 1966", Charles J. Fillmore, pp. 30-64. (Also in General Linguistics 9 (1969), pp. 41-65.)

"Types of Lexical Information", Charles J. Fillmore , pp. 65-103. (Also in Studies in Syntax and Semantics, F. Kiefer, ed., D. Reidel: Dordrecht-Holland (1970), pp. 109-137; and Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, Anthropology, and Psychology, Jacobovits and Steinberg, eds., Cambridge University Press (to appear).)

"'Being' and 'Having' in Estonian", Ilse Lehiste, pp. 104-128. (Also in Foundations of Language 5 (1969), pp. 324-341.)

*No longer available

*No. 3 (June, 1969) (OSU-CISRC-TR-69-4)

"Do from Occur", Gregory Lee, pp. 1-21.
"The Syntax of the Verb 'Happen'," Dale E. Elliott, pp. 22-35.
"Subjects and Agents", Gregory Lee, pp. 36-113.
"Modal Auxiliaries in Infinitive Clauses in English", D. Terence Langendoen, pp. 114-121.
"Some Problems in the Description of English Accentuation", D. Terence Langendoen, pp. 122-142.
"Some Observations Concerning the Third Tone in Latvian", Ilse Lehiste, pp. 143-158.
"On the Syntax and Semantics of English Modals", Shuan-fan Huang, pp. 159-181.

*No. 4 (May, 1970) (OSU-CISRC-TR-70-26)

"Copying and Order-changing Transformations in Modern Greek", Gaberell Drachman, pp. 1-30.
"Subjects, Speakers and Roles", Charles J. Fillmore, pp. 31-63. (Also in Synthese 11 (1970), pp. 3-26.)
"The Deep Structure of Indirect Discourse", Gregory Lee, pp. 64-73.
"A Note on Manner Adverbs", Patricia Lee, pp. 74-84.
"Grammatical Variability and the Difference between Native and Non-native Speakers", Ilse Lehiste, pp. 85-94. (To appear in Proceedings of the 2nd International Congress of Applied Linguistics, Cambridge, England (1969).)
"Temporal Organization of Spoken Language", Ilse Lehiste, pp. 95-114, (Also in Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen, L. L. Hammerich, Roman Jakobson, and Eberhard Zwirner, eds., Copenhagen: Akademisk Forlag (1971) pp. 159-169.)
"More on Nez-Perce: On Alternative Analyses", Arnold M. Zwicky, pp. 115-126. (To appear in IJAL.)
"Greek Variables and the Sanskrit ruki Class", Arnold M. Zwicky, pp. 127-136. (Also in Linguistic Inquiry 1.4 (1970), pp. 549-555.)
"Review of W. F. Mulder, Sets and Relations in Phonology: An Axiomatic Approach to the Description of Speech", Arnold M. Zwicky, pp. 137-141. (To appear in Foundations of Language.)
"A Double Regularity in the Acquisition of English Verb Morphology", Arnold M. Zwicky, pp. 142-148. (To appear in Papers in Linguistics.)
"An Annotated Bibliography on the Acquisition of English Verbal Morphology", Mary Louise Edwards, pp. 149-164.

*No longer available

No. 5 (June, 1969)

Twana Phonology, Gaberell Drachman, pp. 8-286, Ph.D. Dissertation, University of Chicago (1969). [Limited printing: not sent out to everyone on the mailing list.]

*No. 6 (September, 1970) (OSU-CISRC-TR-70-12)

"On Generativity", Charles J. Fillmore, pp. 1-19.
"Relative Clause Structures and Constraints on Types of Complex Sentences", Sandra Annear Thompson, pp. 20-40.
"The Deep Structure of Relative Clauses", Sandra Annear Thompson, pp. 41-58.
"Speech Synthesis Project", David Meltzer, pp. 59-65.
"A Speech Production Model for Synthesis-by-rule", Marcel A. A. Tatham, pp. 66-87.
"Translation of A Model of Speech Perception by Humans, by Bondarko, et al.", Ilse Lehiste, pp. 88-132.

No. 7 (February, 1971) (OSU-CISRC-TR-71-2)

"On Coreferentiality Constraints and Equi-NP-Deletion in English", Alexander Grosu, pp. G1-G111.
Subjects and Agents: II, Gregory Lee, pp. L1-L118, Ph.D. Dissertation, Ohio State University (1970).

No. 8 (June, 1971) (OSU-CISRC-TR-71-7)

The Grammar of Emotive and Exclamatory Sentences in English, Dale Elliott, pp. viii-110.
"Linguistics as Chemistry: The Substance Theory of Semantic Primes", Arnold M. Zwicky, pp. 111-135.
"On Perceptual and Grammatical Constraints Constraints", Alexander Grosu, pp. 136-149.
"On Invited Inferences", Michael Geis and Arnold Zwicky, pp. 150-155.
"Remarks on Directionality", Arnold M. Zwicky, 156-163.
"Evidence", Barry Nobel, pp. 164-172.
"How Come and What For", Arnold M. Zwicky and Ann D. Zwicky, pp. 173-185.
"In a Manner of Speaking", Arnold M. Zwicky, pp. 186-196.

---

*No longer available

Introduction

The papers contained in this issue of Working Papers in Linguistics deal mainly with experimental topics. Units in Speech Perception, by Z. S. Bond, constitutes her dissertation. The next three papers, by L. Shockey, R. Gregorski, and I. Lehiste, deal with various aspects of the temporal structure of spoken language. M. V. Wendell's paper, "Relative Intelligibility of Five Dialects of English", is her undergraduate honors thesis. The volume concludes with three papers devoted to specific languages. Of these, the papers on Hungarian and Estonian are based on experimental techniques; the paper on Latvian and Lithuanian deals with historical phonology. Z. S. Bond's dissertation, I. Lehiste's paper, and the two papers written jointly by L. Shockey, R. Gregorski and I. Lehiste were partly supported by the National Science Foundation under Grant No. GN-534.1. The other papers are published with support from the Graduate School of The Ohio State University.

Units in Speech Perception*

Zinny Sans Bond

---

*Sponsored in part by the National Science Foundation through Grant GN-534.1 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

ACKNOWLEDGMENTS

The help of many people has made this study possible. Above all, I want to thank my adviser, Professor Ilse Lehiste, whose insistence on clear formulation of ideas prevented me from trying many unworkable schemes. I also want to thank the members of my dissertation committee: Professors Catherine Callaghan, Gaberell Drachman, Arnold Zwicky, and, particularly, Neil Johnson. Preston Carmichael, technician, spent many hours helping me design and assemble instrumentation. I sincerely appreciate his help. Tom Whitney gave much assistance in devising and using computer programs for data processing. I thank him for his kind assistance. I also want to thank the Ohio State University Instruction and Research Computer Center for permitting me to use their facilities. I am grateful to the "Phonetics Research Group"--Sara Garnes, Dick Gregorski, Linda Shockey, and Mary Wendell--for listening patiently to my problems and offering many helpful suggestions. I appreciate the help of all the members of the Ohio State University Linguistics Department who kindly served as subjects in many of these experiments. Finally, I want to thank my family for their patience and moral support while I was writing this dissertation.

This work was sponsored in part by the National Science Foundation through Grant GN-534.1 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ii

VITA iii

LIST OF TABLES v

LIST OF FIGURES viii

INTRODUCTION 1

Chapter

I. MODELS OF SPEECH PERCEPTION 4

Behaviorism
Information Theory
Filtering
The Motor Theory of Speech Perception
Analysis by Synthesis
Perceptual Strategies

II. THE PERCEPTION OF SUB-PHONEMIC DIFFERENCES 25

Method
Results
Discussion

III. THE PERCEPTION OF OBSTRUENT CLUSTERS 45

Method
Results
Discussion

IV. SYNTACTIC UNITS IN PERCEPTION 76

Method
Results
Discussion

V. CONCLUSION 91

The Need for Perceptual Units
Implications for Perception Models

BIBLIOGRAPHY 98

LIST OF TABLES

Table | Page
1. Per Cent Correct Identifications 30

2. Consistency of Subjects' Responses 32

3. Subjects' Performance in Relation to Judgments of Ease and Difficulty 33

4. Ease and Difficulty of Word Pairs 34

5. Reaction Time for Subjects with Training in Phonetics for the Pairs Wade/weighed, etc. 35

6. Reaction Time for Subjects with Training in Phonetics for the Pairs Baste/based, etc. 36

7. Reaction Time for Phonetically Untrained Subjects for the Pairs Wade/weighed, etc. 37

8. Reaction Time for Phonetically Untrained Subjects for the Pairs Baste/based, etc. 38

9. Reaction Time to Productions Labeled Consistently. . . . 39

10. Reaction Time to the Mono-morphemic and Bi-morphemic Words Weighed/wade, etc. 41

11. Reaction Time to the Mono-morphemic and Bi-morphemic Words Lacks/lax, etc. 42

12. Reaction Time to the Mono-morphemic and Bi-morphemic Words Missed/mist, etc. 43

13. All Responses--Signal to Noise Ratio: +12 d.b. 53

14. All Responses--Signal to Noise Ratio: 0 d.b. 53

15. All Responses--Signal to Noise Ratio: -6 d.b. 54

16. All Written Responses--Signal to Noise Ratio: +12 d.b. . 55

17. All Written Responses--Signal to Noise Ratio: 0 d.b. . 55

18. All Written Responses--Signal to Noise Ratio: -6 d.b. . 56

Table Page

19. Total Spoken Responses--Signal to Noise Ratio: +12 d.b. . 57

20. Total Spoken Responses--Signal to Noise Ratio: 0 d.b. . 57

21. Total Spoken Responses--Signal to Noise Ratio: -6 d.b. . 58

22. Written Responses for Two-Syllable Words--Signal to Noise Ratio: +12 d.b. 59

23. Written Responses for Two-Syllable Words--Signal to Noise Ratio: 0 d.b. 59

24. Spoken Responses for Two-Syllable Words--Signal to Noise Ratio: +12 d.b. 60

25. Spoken Responses for Two-Syllable Words--Signal to Noise Ratio: 0 d.b. 60

26. Written Responses for Two-Syllable Words--Signal to Noise Ratio: -6 d.b. 61

27. Spoken Responses for Two-Syllable Words--Signal to Noise Ratio: -6 d.b. 62

28. Spoken Responses for [ɪ]--Signal to Noise Ratio: +12 d.b. 63

29. Spoken Responses for [æ]--Signal to Noise Ratio: +12 d.b. 63

30. Spoken Responses for [ʊ] and [oʊ]--Signal to Noise Ratio: +12 d.b. 63

31. Spoken Responses for [ɪ]--Signal to Noise Ratio: 0 d.b. . 64

32. Spoken Responses for [æ]--Signal to Noise Ratio: 0 d.b. . 64

33. Spoken Responses for [ʊ] and [oʊ]--Signal to Noise Ratio: 0 d.b. 64

34. Spoken Responses for [ɪ]--Signal to Noise Ratio: -6 d.b. 65

35. Spoken Responses for [æ]--Signal to Noise Ratio: -6 d.b. 65

36. Spoken Responses for [ʊ] and [oʊ]--Signal to Noise Ratio: -6 d.b. 65

37. Written Responses for [æ]--Signal to Noise Ratio: -6 d.b. 66

38. Written Responses for [ɪ]--Signal to Noise Ratio: -6 d.b. 66

39. Written Responses for [ʊ] and [oʊ]--Signal to Noise Ratio: -6 d.b. 66

Table Page
40. Written Responses for [ɪ]--Signal to Noise Ratio: 0 d.b. . 67

41. Written Responses for [æ]--Signal to Noise Ratio: 0 d.b. 67

42. Written Responses for [ʊ] and [oʊ]--Signal to Noise Ratio: 0 d.b. 67

43. Written Responses for [æ]--Signal to Noise Ratio: +12 d.b. 68

44. Written Responses for [ɪ]--Signal to Noise Ratio: +12 d.b. 68

45. Written Responses for [ʊ] and [oʊ]--Signal to Noise Ratio: +12 d.b. 68

46. Written Responses for Bi-morphemic Words--Signal to Noise Ratio: +12 d.b. 70

47. Written Responses for Bi-morphemic Words--Signal to Noise Ratio: 0 d.b. 70

48. Written Responses for Bi-morphemic Words--Signal to Noise Ratio: -6 d.b. 70

49. Spoken Responses for Bi-morphemic Words--Signal to Noise Ratio: +12 d.b. 71

50. Spoken Responses for Bi-morphemic Words--Signal to Noise Ratio: 0 d.b. 71

51. Spoken Responses for Bi-morphemic Words--Signal to Noise Ratio: -6 d.b. 71

52. Reaction Time for Correct and Incorrect Responses 72

53. Reaction Time to Consonant Clusters 73

54. Mean Reaction Time to Clicks 81

55. Click Localization: Per Cent Correct 88

LIST OF FIGURES

Figure Page

1. Three-stage Mediation-integration Model 8

2. Model of a Closed Cycle Control System for Speaking . . . 13

3. A Model of Speech Communication 14

4. Analysis by Synthesis Model 20

5. Model for the Speech-generating and Speech-perception Process 21

6. Instrumentation for Experiment Testing the Perception of Sub-phonemic Phonetic Differences 28

7. Per Cent Correct Identifications for Each Word Pair . . . 31

8. Instrumentation for Adding Noise to Stimulus Tape . . . 50

9. Instrumentation for "Click" Experiment 80

10. Reaction Time to Clicks in Consonants Preceding Stressed Vowels and to Clicks in Consonants Preceding Unstressed Vowels 82

11. Reaction Time to Clicks in Stressed Vowels and in Unstressed Vowels 83

12. Simple Reaction Time to Click, and Reaction Time to Click in a Constituent Boundary 84

13. Click Localization When the Click Occurs in a Constituent Boundary 85

14. Click Localization 85-87

15. Click Localization in Stressed and Unstressed Vowels . . 88

INTRODUCTION

Speech perception, as a field of empirical investigation, is very much involved with linguistics: a model of speech perception is crucially dependent on a model of language, since the model of language tells the perception theorist what it is that the listener has to perceive.

Thus, historically, there has been a tendency for models of speech perception to be related to the current linguistic models of language. The early models of speech perception are not specific enough, by current standards, simply because the model of language that the theorist was dealing with was not a very complex model--language was conceived to be something like a series of words strung together.

As more complicated and more precise linguistic models become current, the theorizing about speech perception also became more precise and more experimentally oriented. Thus, structural linguistics of the 1940's and 1950's led to experimental work which assumed that the phoneme, or some unit very much like a phoneme, was the perceptual unit in phonology. The problem in understanding speech perception was then seen as discovering how a listener can 'translate' or 'decode' a continuous acoustic signal into discrete phonemes. And, though alternative suggestions have been made, most theorists still assume that the incoming speech signal is represented in some phoneme-like units as the first step in speech perception.

Experimental work on higher-level perceptual units, related to the syntactic structure of a sentence, has begun quite recently. Some early theorists have advanced ideas of what is involved in understanding sentences, but, again, the work could not lead to any precise theoretical formulations until a fairly adequate theory of syntax became available; thus, almost all empirical studies involving the perception of syntactic units assume that the syntactic relationships described in transformational grammar are involved in speech perception at some level. However, the experiments have tended not to separate perceptual effects from memory effects; and there is no agreement--such as implicitly exists in theories of the perception of phonological segments--whether there are some syntactic units involved in perception and, if so, what these units are.

Generative phonology, which does not assume any unit equivalent to the traditional phoneme, has not so far led to any experimental work on speech perception, though it is intimately related to models of speech perception involving analysis-by-synthesis.

In this study, the attempt is made to examine some units that function in speech perception. The first chapter contains a survey of models that have been proposed to account for speech perception. The survey includes some models because of the historical background they provide, even though the models make no specific predictions about units in speech perception. More recent models make certain predictions about perceptual units, and these will be pointed out when the theoretical implications of the perceptual models are discussed.

Three experiments are reported. The first experiment involves a subject's ability to make use of sub-phonemic phonetic differences.

Subjects are asked to identify productions of mono-morphemic and bimorphemic words of identical phonemic shape, e.g., lax vs. lacks. The purpose of the experiment is two-fold: to determine what a 'baseline' for perception is--what is the least amount of phonetic difference that can be used for linguistic purposes--and to determine if the traditional phoneme, which is often accepted as the perceptual unit, defines a lower limit below which a listener can not make use of phonetic differences.

The second experiment involves the perception of obstruent clusters. Subjects are asked to identify words with reversible obstruent clusters, such as task vs. tax, in the presence of noise. The purpose of the experiment is to determine whether consonant clusters are coded 'phoneme-by-phoneme', as the traditional assumptions would imply, or if subjects employ some alternative perceptual mechanisms.

The third experiment seeks to determine perceptual units in syntax. Subjects are asked to respond, by pressing a button, when they hear a 'click' in a sentence. From reaction time to the click, the effects of a phonologically defined phrase on perceptual segmentation can be determined.

Finally, the implications of the experimental studies to models of speech perception are discussed.

CHAPTER ONE

MODELS OF SPEECH PERCEPTION

The purpose of this chapter is to provide some historical background and to present the current ideas of theorists attempting to account for speech perception. Not all of the models that will be discussed in this chapter make specific predictions about what units are involved in speech perception, but they are included simply because many are interesting in themselves or for historical reasons.

No attempt will be made to evaluate the adequacy of any of these models in this chapter. Rather, the models that still hold promise will be discussed in the last chapter in terms of the theoretical implications of the empirical studies reported in this work.

Models of speech perception have been classified under the following headings: behavioristic models, information theory models, motor theories, analysis by synthesis models, models proposing 'filtering' as a primary device, and models depending on perceptual strategies.

Behaviorism

There is a long behaviorist tradition of theories of speech perception. Appropriately enough, it begins with J. B. Watson (1930). Watson's general behaviorist position is well known, and his views of language--not developed in any great detail--follow from it clearly. Since he refuses to postulate any "mentalistic constructs," he discusses language in observable, physicalistic terms. Language is simply a "manipulative habit of the vocal tract" (Watson, p. 225). When a person learns to speak, he develops a conditioned response--some movement of the vocal tract--for every object and situation in his external environment. These conditioned responses are equivalent to words. Such internalized kinaesthetic responses can call out further reponses in the same way as the objects for which they serve as substitutes do; because of these kinaesthetic verbal substitutes, a person carries the world around with him; he can manipulate the world (think) by means of series of motor responses.

Sentences, and other language sequences, are accounted for by the following example: a child hears the bed time prayer "Now I lay me down to sleep..." The first few times he hears it, the first word of the sentence, "now," makes the child produce the motor response which is his internal equivalent of "now;" similarly "I" leads to internalized "I," etc. After repeated experiences, the motor response "now" will lead directly to the motor response "I," with no necessary intervening step. At this point, the child has learned the sentence. Spontaneous speech, Watson believes, follows essentially the same principles: some stimulus touches off old verbal organization.

Speech perception offers no particular difficulty: the incoming stimulus makes the listener form the equivalent kinaesthetic-motor responses. Watson, therefore, is postulating a simple motor theory of speech perception, involving incipient muscle activity.

In Language, Bloomfield (1933) offers a much more sophisticated analysis of language, but his outlook is essentially behavioristic.

Bloomfield analyzes an event involving speech by means of a little scene with two characters, Jack and Jill. Externally, the action is quite simple: Jack and Jill are walking along a road; Jill makes a series of noises with her vocal tract; Jack climbs a fence, and brings Jill an apple from a nearby tree.

Looking at the scene more analytically, there are a number of practical events preceding the act of speech. These practical events are quite complex, but taken together, they can be considered as a stimulus for Jill. As a speaking human, Jill has a choice: she can make a direct response (go get the apple), or she can make a linguistic substitute response (ask Jack for the apple). For Jack, the speech is a substitute linguistic stimulus, which makes him produce a particular response.

Essentially, speech enables stimuli and responses to occur in different individuals, as indicated in the following diagram:

$$S \rightarrow r \cdots\cdots s \rightarrow R$$

Bloomfield is not very specific in discussing what is involved in Jack's reception of the message. In relation to phonology, Bloomfield argues that speakers of a language habitually and conventionally discriminate some features of sound and ignore others; presumably, then, there are distinctive properties of sound to which Jack is sensitive. These encode the message.

The behaviorist tradition is carried on in the 1950's by the psychologists B. F. Skinner (1957), O. H. Mowren (1954), and C. E. Osgood (1963).

Mowrer does not offer a complete theory of language, but an analysis of declarative sentences in stimulus-response (henceforth S-R) terminology. Essentially, he suggests that a sentence is an arrangement for conditioning the meaning reaction produced by the predicate to the stimulation aroused by the meaning reaction elicited by the subject. In other words, a subject-predicate sentence is to be considered a conditioning device.

The conditioning device operates in the following way. When the listener hears any word in his vocabulary, there is aroused in him a unique "meaning response." When he hears a sentence, for example, "Tom is a thief," first there is aroused in the listener a "meaning response" which is his internal representation of the word "Tom" as well as of the physical Tom. Then, because a sentence is a conditioning device, to this "meaning response" is added the "meaning response" of "thief." As a consequence, the listener comes to respond differently to the physical Tom; he will avoid him, perhaps, and not lend him money. In short, he will treat Tom as a thief.

One of the most thorough attempts to explain language behavior in S-R terms is B. F. Skinner's book Verbal Behavior (1958). Skinner declines to speculate about non-observable language phenomena; rather, he sees the task of the science of verbal behavior to determine the laws governing verbal behavior. These laws concern the predictability and control of particular verbal responses. That is, the task is accomplished when it is possible to predict what a person will say.

Because of this goal, and because he rejects non-observables, Skinner has little to say about internal phenomena such as perception. He does offer a few suggestions. First, Skinner defines a unit of verbal behavior as anything that is under the independent control of a manipulable (stimulus) variable. This unit can be as large as a whole phrase, such as "How are you?", or as small as a change in fundamental frequency, used to ask a question. In order for language to function at all, these units must lead to different responses by listeners. Secondly, Skinner points out that at any time in sequential verbal behavior, e.g. sentences, what has been said before sharply limits what will be said next: there is redundancy in language. Presumably, the listener can also take advantage of such redundancy.

But Skinner does not attempt to present any theory of speech perception; the few suggestions that he makes do not detract from his basic assumption that perception can not be separated from responses in any meaningful way.

C. E. Osgood also offers a behavioristic theory of speech (Osgood, 1963), which he calls a three-stage mediation model. Unlike Skinner, Osgood is quite ready to postulate mechanisms internal to the speaker and listener. Rather than being concerned only with observable stimuli and responses, Osgood wants to fill the "black box" of the organism with intervening S-R constructs. Osgood's three-stage model is represented below.

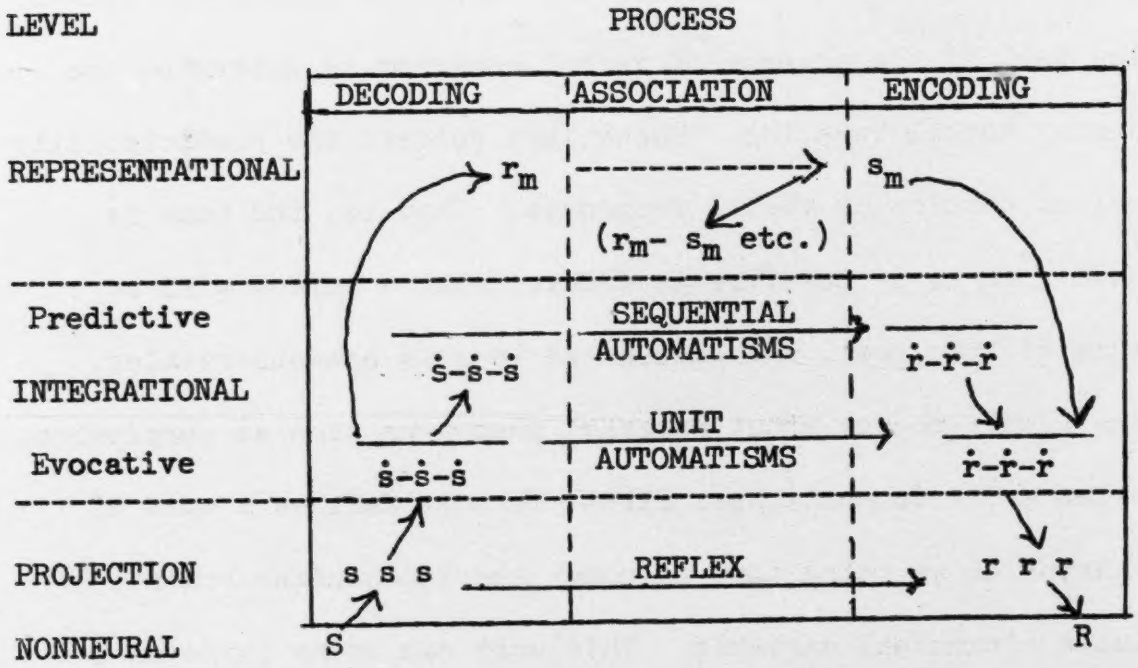

LEVEL
PROCESS
DECODING
ASSOCIATION
ENCODING
REPRESENTATIONAL
r_m
s_m
(r_m- s_m etc.)
Predictive
SEQUENTIAL AUTOMATISMS
$\dot{s}$-$\dot{s}$-$\dot{s}$
$\dot{r}$-$\dot{r}$-$\dot{r}$
INTEGRATIONAL
UNIT AUTOMATISMS
Evocative
$\dot{s}$-$\dot{s}$-$\dot{s}$
$\dot{r}$-$\dot{r}$-$\dot{r}$
PROJECTION
s s s
REFLEX
r r r
NONNEURAL
S
R

Fig. 1. Three-stage mediation-integration model.

(by permission of Charles E. Osgood)

Osgood's model differs from Skinnerian S-R models in two ways. First, Osgood postulated mediating responses (r_m). These internal r_m's are a fractional, easily differentiable, part of an original overt response. Since the original response was elicited by some stimulus, the fractional r_m becomes an internal representation of the stimulus. The internal r_m's, in turn, can lead to various instrumental acts. Essentially, Osgood hopes to account for meaning by these internal representations. These internal representations, however, are quite complex; basically, Osgood holds that words are coded by means of a simultaneous bundle of semantic features (Osgood, 1963).

Secondly, Osgood postulates stimulus integration (S-S learning) and response integration (R-R learning) to account for the perceptual and motor complexity found in speech. He argues that, in perception, the greater the frequency with which stimulus events have been paired in the input experience of the organism, the greater will be the tendency for their central neural correlates to activate each other. In other words, a partial sensory input will become adequate to trigger the whole; it will lead to what the Gestalt psychologists called "closure."

This closure principle can only operate if there are perceptual units which function as wholes. These units must meet three criteria: they must be highly redundant, they must be fairly frequent in occurrence, and they must not exceed certain temporal limits. The most likely perceptual units are words.

In perceiving a sentence, the phonetic information is adequate to trigger the phonological representation of a particular word, e.g.

play. The context of the sentence then determines the semantic interpretation of the word. Given, for example, the sentence "The play got rave reviews," the word *play* will be interpreted as a noun on the basis of the frame Determiner _____ verb. The word *review* will eliminate the interpretation of *play* in the sense of gambling. On the basis of such linguistic information and on the basis of non-linguistic context, the listener will arrive at the intended message.

More recently, psychologists, even though they may consider themselves behaviorists, have broken away from S-R formulations altogether.

In his very interesting book, *The Senses Considered as Perceptual Systems*, James J. Gibson (1966) emphasizes the information contained in stimulation, rather than the discrete responses of separate sensory systems. Therefore, he rejects the traditional decomposition of a complex sound into a combination of pitch, duration, and loudness specifications in order to describe the stimulus. He considers it a better approach to look for higher-order variables characteristic of the stimulus:

> "In meaningful sounds, these variables can be combined to yield higher-order variables of staggering complexity. But these mathematical complexities seem nevertheless to be the simplicities of auditory informaiton, and it is just these variables that are distinguished naturally by an auditory system." (p. 87).

In other words, it is a mistake to think that the perceptual system "builds up" complex stimuli from simple components; rather, complex stimuli are responded to directly.

The higher-order variables have not been studied for most types of meaningful sound, but there have been a few attempts to study such variables in the acoustic speech signal. According to Gibson, frequency ratios and the relational patterns of frequencies are the invariants provided by the speech signal.

The pick-up of phonemes is a direct one-stage process; however, the apprehension of things referred to--a semantic decoding of the speech signal--is a two-stage process since not only the speech sounds but what they stand for have to be apprehended. "The acoustic sounds of speech specify the consonants, vowels, syllables, and words of speech; the parts of speech in turn specify something else." (p. 91).

The structure of speech can be analyzed at various levels, hierarchically organized, and each level has some unit appropriate to it: at each level, there is an appropriate stimulus unit for the perceptual system.

Information Theory

During the 1950's, information theory provided conceptual structures by which all types of communication--defined as the transmission of information--could be analyzed. Theorists concerned with speech also tried to apply the concepts of information theory to their field, and developed models of speech communication. These speech communication models discussed both a speaker and a hearer, but tended to emphasize the former. Many models of the speech communication system were proposed; these are summarized by Grant Fairbanks (1954), who also presents one of the most detailed analyses of speech from this point of view. However, most of his discussion concerns speech production. Perception is discussed almost exclusively in terms of its role in feedback: the speaker monitors his own output and changes his output when it does not meet the criteria set by the input to the speech systems.

Fairbanks' model is reproduced in Fig. 2. Essentially, the model offers the following analysis of speech production: an input signal to the speech mechanisms results in some output; this output is compared with the stored input; if the output has not yet reached the target specified by the input, an error signal is sent out to adjust the output.

There are several interesting points concerning the speech model. First, Fairbanks postulates a "unit of control." Although he does not go into detail, he suggests that the unit of speech control is not to be identified with any currently recognized phonetic unit; rather, the unit of speech control is a "semi-periodic, relatively long, articulatory cycle" (p. 138). Secondly, the model implies that certain steady-state outputs are the goals of the speech mechanism and that transitions are only by-products. In Fairbanks' words:

> "It is to be emphasized that the steady states are the primary objectives, the targets. The transitions are useful incidents on the way to the targets. The roles of both are probably very analogous when the dynamic speech output is perceived by an independent listeners." (p. 139)

Fairbanks has little to say about speech perception directly. Presumably, perception follows the path described for feedback. Whether the message is analyzed directly or whether it is compared in the comparator with a possible message--as in motor theories of speech perception--is not specified in Fairbanks' model.

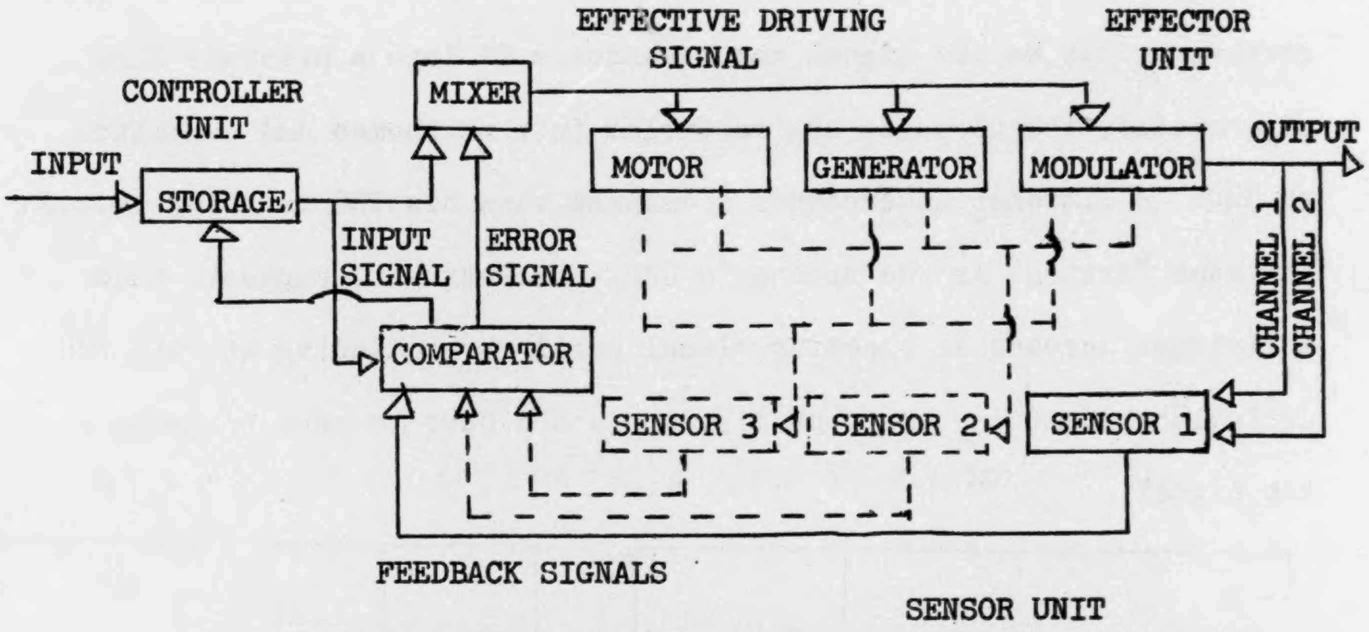

EFFECTIVE DRIVING SIGNAL
EFFECTOR UNIT
CONTROLLER UNIT
MIXER
INPUT
STORAGE
MOTOR
GENERATOR
MODULATOR
OUTPUT
INPUT SIGNAL
ERROR SIGNAL
COMPARATOR
CHANNEL 1
CHANNEL 2
SENSOR 3
SENSOR 2
SENSOR 1
FEEDBACK SIGNALS
SENSOR UNIT

Fig. 2. Model of a closed cycle control system for speaking. (Grant Fairbanks, "A Theory of the Speech Mechanism as a Servo-System." Journal of Speech and Hearing Disorders 19 (1954). By permission of the American Speech and Hearing Association).

Although it uses concepts from information theory, Hockett's model of speech communication (1956) is much more linguistic in orientation than Fairbanks' model, at least in the sense that linguistic terminology is applied to various processes. However, Hockett cautions that the 'phoneme' and 'morpheme' of internal circuitry are not to be strictly equated with the phoneme and morpheme of linguistics.

Hockett's model (Fig. 3) represents the internal mechanisms necessary for Jill to communicate with Jack. First, a sequence of morphemes is emitted by GHQ (grammatical headquarters); then the morphemes are recoded into a discrete flow of phonemes by morphophonemic processes. Finally, the phonemes become a continuous speech signal in the "speech transmitter." The speaker monitors his own speech signal, but he does not use feedback to adjust the output continuously.

The listener uses the same communications system, but the speech receiver sends the signal through in the other direction; the speech receiver picks up the signal and transduces it into a discrete flow of phonemes; the phonemes are assembled into morphemes and submitted to GHQ. A listener understands a message when his GHQ is going through the same "states" as the speaker's GHQ. Hockett also suggests that a listener decodes an incoming signal partly by comparing it with the articulatory motions that the listener would have to make to produce the signal.

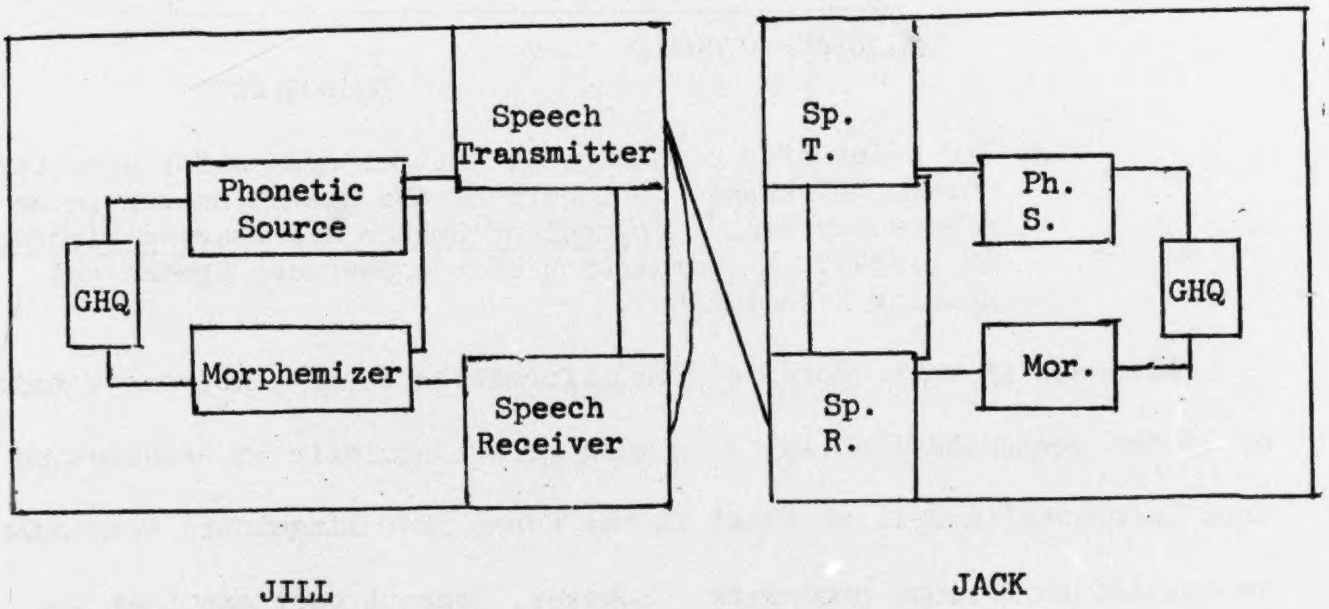


Fig. 3. A model of speech communication. (Charles Hockett, A Manual of Phonology, 1955, by permission of Indiana University Publications in Anthropology and Linguistics and Prof. Charles F. Hockett.)

Filtering

In his article "On the Process of Speech Perception," J. C. R. Licklider (1952) analyzes the process of speech perception into three main operations: translation of the speech signal into a form suitable for the nervous system, identification of speech elements, and comprehension of meaning.

The first process is performed by the cochlea; the signal is mechanically analyzed in terms of frequency and intensity in such a way that the output is somewhat similar to a sound spectrogram. However, since the frequency analysis of the cochlea is not very selective, the signal is sharpened further up the auditory pathways. Thus, the input to the perceptual mechanism consists of a sharpened frequency analysis of the acoustic signal, coded in terms of origin on the cochlea, and intensity, coded in terms of density of discharge. Furthermore, there is a representation of the fundamental frequencies of the periodic components of the acoustic signal.

The second process, identification of speech elements, could be performed by one of two mechanisms, a correlator or a filter. A correlator is essentially a device for matching the incoming signal against an internally stored representation (or a representation created by rules). A filter, on the other hand, has the required patterns built into its structure; the identification of the incoming signal is made on the basis of which filter the signal passes through most successfully. Although the choice is tentative, Licklider favors the filter model as the device which identifies speech elements.

Comprehension, on the other hand, can best be explained as an active process. Therefore, Licklider argues that comprehension of meaning involves matching the input to a set of internal patterns. Although he does not say this, Licklider would probably maintain that these patterns are generated as needed.

Licklider's model, therefore, is very much like analysis-by-synthesis for the processing of sentences. For smaller units, however, Licklider prefers the more direct analysis provided by filtering.

A "filtering" theory, differing in interesting ways from Licklider's, has been recently developed by Wayne A. Wickelgren (1969a, 1969b). Previous theories have assumed that, no matter how speech is processed, the phoneme is the primary unit of coding in perception. Wickelgren proposes a theory in which the perception and production of speech is coded in some unit that is more closely related to the traditional allophone. He calls this theory context-sensitive coding.

> "I define a context-sensitive code for words to consist of an unordered set of symbols for every word, where each symbol restricts the choice of its left and right neighbors sufficiently to determine them uniquely out of the unordered set for any given word. In this case, the unordered set, in conjunction with the dependency rules, contains all the information necessary to reconstruct a unique ordering of the symbols for each word." (1969b, p. 86)

In speech perception, context-sensitive coding would work in the following way. Each context-sensitive allophone of the language would have a unique internal representative. This internal representative would be activated by some conjunction of acoustic features, occurring over a period of time as long as a few hundred milliseconds. All allophone representatives would be examining the acoustic input in parallel, but only a few would be activated in response to the input. After the set of allophones has been determined, the word representative which is most closely associated with the set of allophones can be selected.

Wickelgren claims that his theory eliminates two of the major problems associated with perception models which postulate phonemes as the basic units: first, there is no need to segment the acoustic wave form; second, it is more likely--although the evidence is not in--that there is invariance in the acoustic signal for allophones.

The model of speech perception proposed by L. V. Bondarko and others (Bondarko et al., 1970) is designed to account for the set of operations that transform an acoustic speech signal into a sequence of words. Each word in the output would have associated with it a set of lexical and grammatical features which would be employed in understanding the message.

The model consists of hierarchically-arranged processes. At each level, there is a perceptual procedure, decision making, and a procedure for assigning a certain reliability to the decision. If no decision can be made with a threshold degree of reliability, the level outputs several possible interpretations of the input signal, and the final decision is postponed. The final decision may not be made, in fact, until the last stage--the recognition of the meaning of the utterance.

The first stage of the perceptual process is auditory analysis. The output of the cochlea is described in the set of parameters that are relevant in the perception of speech. The output of the auditory analysis is then classified into phonemes (a phoneme is defined as the subjective image employed by the brain of the listener in the process of speech recognition (p. 114); thus it is not strictly equivalent to the linguistic phoneme). Information distributed over an open syllable is employed in this classification process. At the next level, the string of phonemes is segmented, taking stress into account. Then the segmented string is interpreted as a sequence of words.

The Motor Theory of Speech Perception

Although motor theories of speech perception have been advanced by quite a number of theorists, the most explicit and reasoned statement of the motor theory has been formulated by workers at Haskins Laboratories, namely F. S. Cooper, A. M. Liberman, D. P. Shankweiler, and others. For example, in an early discussion of some of their results (Cooper et al., 1952), the Haskins group advanced the motor theory.

The research as Haskins began with a search for invariants in speech--"A one-to-one correspondence between something half-hidden in the spectrogram and the successive phonemes of the message." (Cooper et al., 1952, p. 604). However, no acoustic invariant could be found for the individual phonemes. In fact, Cooper suggests that the perceived similarities and differences between speech sounds may correspond more closely to the similarities and differences in articulation than to the acoustic signal. As evidence for the simpler relation of perception and articulation, Cooper cites the complex relationship of the frequency of the burst of a stop consonant to the point of articulation: a burst of 1440 cps. is heard as /p/ before /i/ but as /k/ before /a/; conversely, bursts at different frequencies can be heard as the same consonant.

In connection with further work with synthetic speech, the Haskins group advanced the notion of categorial perception: perception of phonemes is different from perception of non-speech stimuli in that listeners can discriminate very little better than they can identify absolutely. An acoustic continuum is categorized into phonemes by listeners but a comparable non-speech continuum is not. Furthermore, listeners show discrimination peaks at phoneme boundaries when the stimulus is speech, but no such peaks in discrimination appear when the stimulus is a comparable non-speech continuum (Liberman, Harris, Kinney, and Lane, 1957). These results, which are typically most clear-cut for stop consonants, are readily explained by the motor theory. It is argued that the gesture used in speech production is essentially invariant for the phoneme; therefore, perception is also invariant and categorial.

In their most detailed explication of the motor theory (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967), the Haskins group recapitulates the many arguments advanced for the motor theory and also specifies at what "level" production is made use of in perception. In their earlier work, the assumption was made that the production invariants were "motor commands" which were identical for each production of a given phoneme. In their latest statement, the idea of motor commands is retained and the theory is extended to higher-level neural signals which stand in a one-to-one relationship with other segments of the language:

> "In phoneme perception...the invariant is found far down in the neuromotor system, at the level of the commands to the muscles. Perception by morphophonemic, morphemic, and syntactic rules of the language would engage the encoding process at higher levels." (p. 454)

In this form, the motor theory becomes equivalent to analysis-by-synthesis, a theory of speech perception dependent on the use of rules in just such a way.

Analysis by Synthesis

Essentially, analysis by synthesis is a model of perception that depends on matching the incoming stimulus to an internally-generated pattern. When the internal pattern matches the stimulus, perception has been successful. As a model for speech perception, analysis by synthesis has been extensively developed by Morris Halle and Kenneth N. Stevens.

An early version of the model (Halle and Stevens, 1964) is diagrammed in Fig. 4.

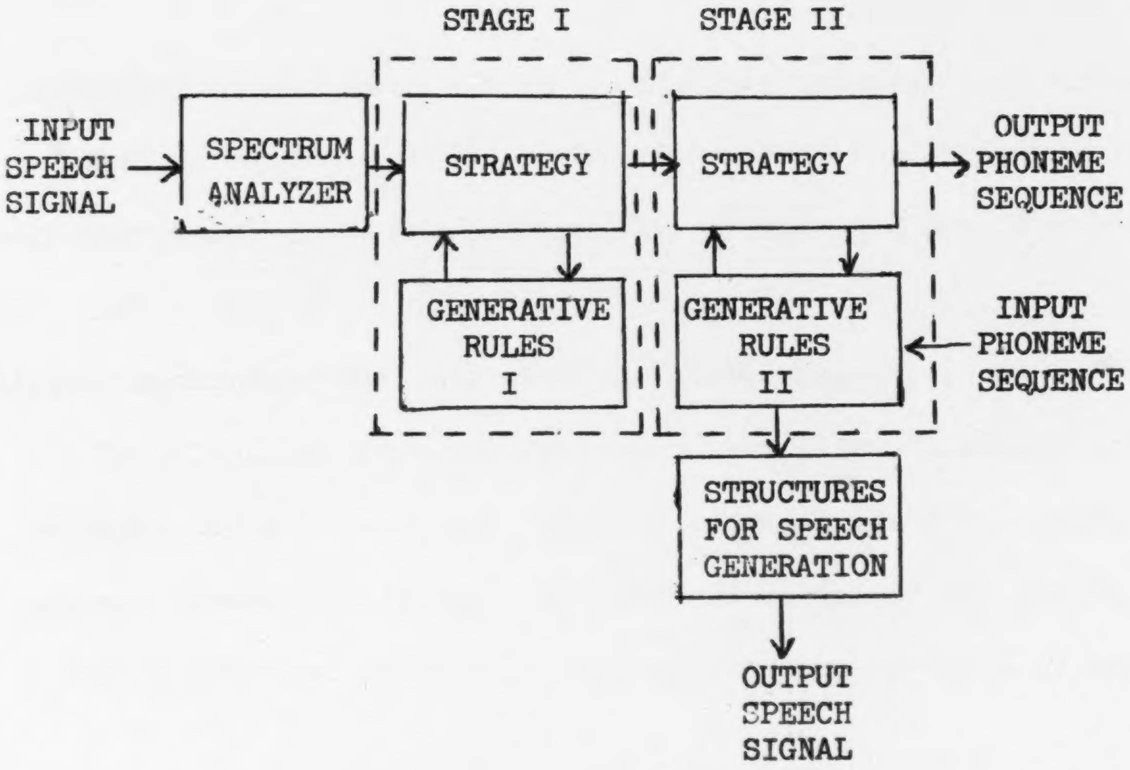

STAGE I
STAGE II
INPUT SPEECH SIGNAL
SPECTRUM ANALYZER
STRATEGY
STRATEGY
OUTPUT PHONEME SEQUENCE
GENERATIVE RULES I
GENERATIVE RULES II
INPUT PHONEME SEQUENCE
STRUCTURES FOR SPEECH GENERATION
OUTPUT SPEECH SIGNAL

Fig. 4. Analysis by Synthesis model.
(Morris Halle and Kenneth N. Stevens, "Speech Recognition: a Model and a Program for Research," in The Structure of Language, ed. by Jerry A. Fodor and Jerrold G. Katz, 1964, by permission of Prentice-Hall).

The model depends on two analysis-by-synthesis loops. After a spectrum analysis, which in large part is a result of cochlear action, the first analysis-by-synthesis loop reduces the spectral representation of the acoustic input to a set of phonetic parameters. This is accomplished by matching the incoming spectrum to a spectrum produced by an internal synthesizer which has the ability to compute spectra when given phonetic parameters. In the second analysis-by-synthesis loop, the phonetic parameters are transformed to a sequence of phonemes. The second loop uses the generative rules that must also be employed in speech production--rules that transform phonemes to phonetic parameters.

In a more recent statement of analysis-by-synthesis (Stevens and Halle, 1965), the analysis-by-synthesis model is integrated with linguistic concepts. The model is represented in Fig. 5.

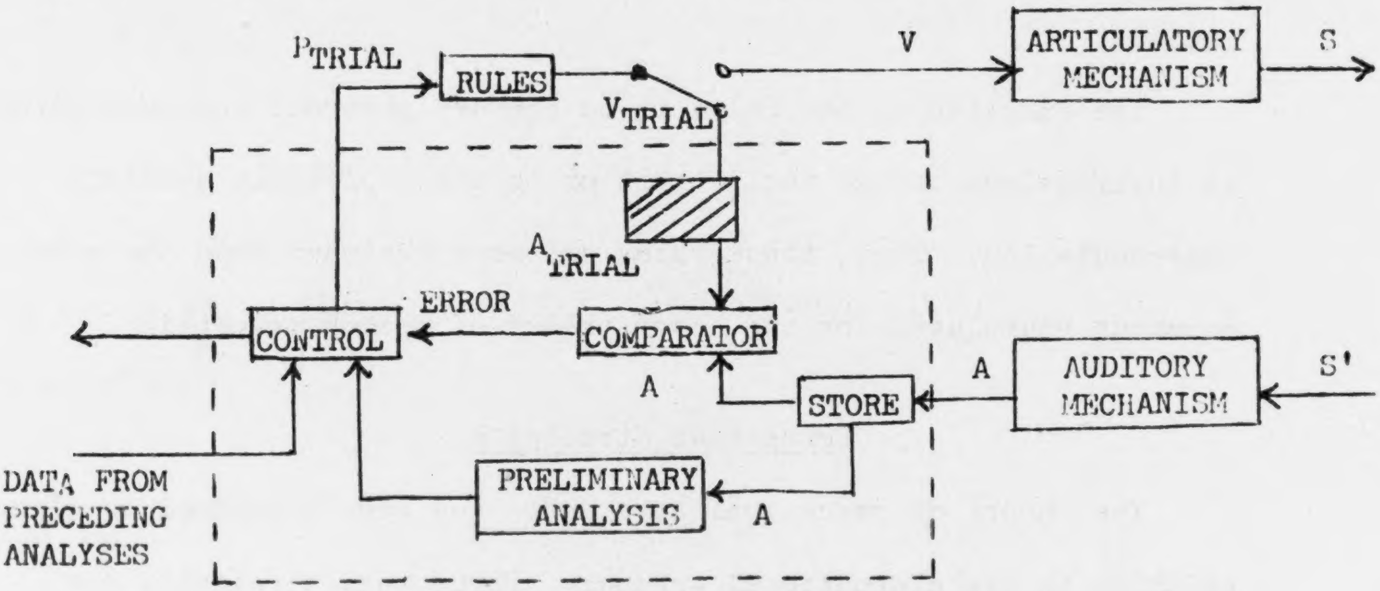

P_{TRIAL}
RULES
V_{TRIAL}
V
ARTICULATORY MECHANISM
S
A_{TRIAL}
ERROR
CONTROL
COMPARATOR
A
STORE
A
AUDITORY MECHANISM
S'
DATA FROM PRECEDING ANALYSES
PRELIMINARY ANALYSIS
A

Fig. 5. Model for the speech-generating and speech-perception process. The dashed line encloses components of a hypothetical analysis-by-synthesis scheme for speech perception. (K. N. Stevens and M. Halle, "Remarks on Analysis by Synthesis and Distinctive Features," in Models for the Perception of Speech and Visual Form, 1965, by permission of M.I.T. Press.)

This model also claims that the mechanism employed in speech production is the same as the mechanism used in speech perception. Furthermore, the model employs abstract representations of words, coded in terms of distinctive features, and phonological rules, apparently identical to the rules found in the phonological component of a generative grammar.

The model operates in the following fashion. The auditory pattern derived from the acoustic input undergoes preliminary analysis; the exact nature of preliminary analysis is not specified in this model. On the basis of the preliminary analysis and contextual information, a hypothesis is made concerning the abstract representation of the utterance. The proposed abstract representation is converted to an equivalent auditory pattern and compared with the pattern under analysis. If there is agreement, then the hypothesized abstract representation is judged to be correct, and processing at more abstract levels can proceed.

The function of the rules is to convert abstract representations to instructions to the vocal tract or to the equivalent auditory representation. Thus, these rules are more abstract than the motor commands postulated for the motor theory of speech perception.

Perceptual Strategies

The theory of perceptual strategies has been developed in close relation to transformational grammar. Perceptual strategies are techniques used by listeners to arrive at a segmentation of a sentence into deep structure units and to assign the proper grammatical function to each component. The theory is the result of research by M. Garrett, J. A. Fodor, and Thomas Bever. At the present, it is in a much more fluid state than the other theories discussed so far, so it seems appropriate to discuss the development of the theory, as well as its current status.

The early statements of the theory (Fodor and Bever, 1965; Garrett, Bever, and Fodor, 1966) were based on the phenomenon of click localization: when presented with a sentence with a superimposed click, the subject locates the click toward the nearest constituent boundary. Furthermore, subjects localize clicks correctly primarily when they occur on a constituent boundary. This phenomenon is interpreted to mean that surface structure constituents form perceptual units, tending to resist interruption by extraneous material.

In later work, more detailed analysis of perceptual strategies followed. Fodor, Garrett, and Bever (Fodor and Garrett, 1967; Fodor, Garrett, and Bever, 1968) suggest that information about the properties of specific lexical items is employed by listeners. The listener selects the verb of the sentence and classifies it according to the possible deep structure configurations it can occur with; then the listener checks all these possible deep structure configurations to see if the surface structure he is presented with is a possible transformational version of the deep structure. In this process of selecting possible deep structures, the subject takes advantage of surface structure markers; for example, "to" implies that the verb must be able to take a "for...to" complementizer.

Later work also indicated that surface structure constituents were not directly related to perception (Bever, Lackner, and Kirk, 1969). Rather, the units of perception seem to be deep structure units.

The current status of the theory of perceptual strategies, as well as a summary of relevant research, has been presented by Bever (1970). In this article, Bever rejects the theory of derivational complexity. This theory claims that the perceptual complexity of a sentence is directly related to the number of transformations involved in its derivation. (A theory of analysis-by-synthesis at a syntactic level would imply derivational complexity.) But Bever finds that, in many cases, transformations are not related to perceptual complexity. First, transformational rules that delete structure do not add complexity; second, certain reordering transformations may even simplify perception. For example, (1) is no more complex--and may even be simpler--than (2);

(1) It amazed Bill that John left early.

(2) That John left early amazed Bill.

Bever then proceeds to discuss several perceptual strategies employed by listeners. Some of these are the following.

a. When faced with a sentence, the listener isolates those adjacent phrases of surface structure which could correspond to a sentence in deep structure. The listener accomplishes this by segmenting together items that could be related as "actor, action, object...modifier."

b. Unless there is information to the contrary, the first noun...verb clause is treated as the main clause.

c. Constructions are related internally according to semantic constraints. Essentially, the listener selects the most likely semantic organization.

d. Any Noun-Verb-Noun sequence that is potentially a unit corresponds to "actor, action, object."

e. The special properties of function words and verbs are employed.

There is no need to give a complete list of proposed perceptual strategies, since all of them are proposed more or less tentatively. The general thrust of the theory, however, is this: to integrate perceptual strategies that are discovered to be applicable in language with other perceptual and cognitive processes, and to determine how language is related to other human cognitive abilities.

CHAPTER TWO

THE PERCEPTION OF SUB-PHONEMIC PHONETIC DIFFERENCES

In the models of speech perception discussed in the preceding chapter, it has been implicitly assumed that phonetic differences that are less than phonemic can have no linguistic significance, and that such differences can not be of any use to the listener. ("Phonemic" is to be understood here as "reliably signaling a difference in meaning.") This assumption follows directly from the traditional notion of a phoneme as a functional unit, distinct from all other such units. This view is also implicit in the notion of "categorial perception of phonemes" recently advaneed by workers at Haskins Laboratories (Studdert-Kennedy, Liberman, Harris, and Cooper, 1970). On the other hand, phoneticians can develop an ability to notice small phonetic differences. And even ordinary listeners are sensitive to non-linguistic information that may be carried by sub-phonemic differences; for example, in identifying a particular speaker, sub-phonemic information is employed. However, speaker identification judgments are not linguistic and may be based on a great deal more information than on the fine phonetic details of an utterance.

In order to establish a "baseline" for perceptual units, it would be helpful to determine exactly how much use a subject can make of non-phonemic phonetic differences for linguistic judgments.

A preliminary study related to this question was conducted by D. B. Fry (1968). Fry found that he was able to identify productions of the two words lax and lacks with no contextual information provided. The experiment was conducted in the following way: Fry prepared a tape by splicing copies of one production of lax and one production of lacks in random order. He then listened to the tape, and, after hearing each word, he pushed a button to identify it. Fry obtained both identification scores and reaction time to the two words. He found, to his surprise, that he could identify the utterances correctly 96 times out of 100 (a statistically significant result). Furthermore, he found that the reaction time to lacks was faster than to lax, although the difference was not statistically significant.

Fry's study is quite tentative, so it is not proper to draw a generalization from it. Fry tested only one subject, himself, and only one supposedly-homophonous word pair. There are a number of possible explanations of the results that do not imply that listeners are generally aware of sub-phonemic differences. First, Fry is a very fine phonetician; therefore, he may be sensitive to distinctions which completely escape the ordinary listener. Second, he may have, by chance, tested very distinctive productions of the two words; ordinarily, the two words may not be nearly so distinctive. Finally, it may be that some error in one or the other of the two words made them distinctive but not in a linguistic sense--there may have been some extraneous noise on the original recording of the utterance.

However, Fry's finding, if it reflects a general listener ability, has considerable implications for theories of speech perception. Therefore, it seemed desirable to replicate Fry's experiment with control over the variables mentioned above.

Method

Stimuli: Ten pairs of words were selected, each pair consisting of one monomorphemic and one bi-morphemic word of the same phonemic shape. Each pair of words composed a sub-list; within the sub-list, the two words were recorded in random order, each word appearing ten times. Each sub-list was introduced by two sentences in which the two words to be tested appeared in context. The following word pairs were tested: wade/weighed, hose/hoes, bard/barred, pact/packed, lax/lacks, baste/based, adds/adze, mist/missed, laps/lapse, and guest/guessed. The speaker was a male graduate student, a speaker of General American, whose home is in Connecticut.

The following procedure was employed to record the stimulus tape: for each production of each word to be recorded, the speaker was presented with a sketch picturing an activity suggestive of the word; underneath the sketch was a sentence employing the word, and descriptive of the sketch. The speaker was certain that under these circumstances he could produce the "correct word."

Two stimulus tapes were recorded; the second tape was a counterbalanced version of the first tape. On both tapes, words within lists were separated by five seconds; sub-lists were separated by ten seconds. Both tapes were recorded in a sound-proof recording booth, on an Ampex 350 tape recorder, at 7 1/2 i.p.s.

Subjects: Two groups of subjects participated in the experiment: 17 undergraduate students with no training in phonetics, and 12 graduate students in an introductory or advanced phonetics class.

The subjects were informed that the purpose of the experiment was to determine how quickly and how accurately people could identify words that sound very much the same. The subjects were instructed to respond as quickly as possible and to guess if they did not know which word they heard.

Procedure: The instrumentation is described in the accompanying diagram (Fig. 6).

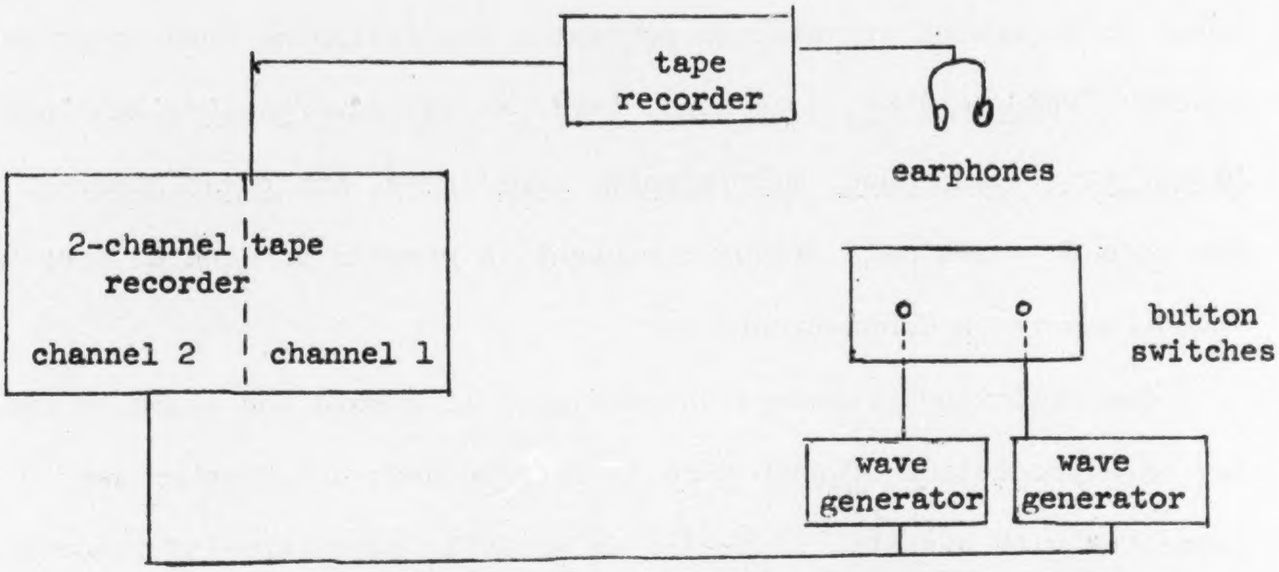


Fig. 6. Instrumentation for experiment testing the perception of sub-phonemic phonetic differences.

Each subject listened to the stimulus tape over earphones; he responded to each word by pushing one of two buttons, which were labeled, to identify which word he heard. The buttons were connected to two signal generators, one generating a sine wave, the other a square wave. Both the stimulus tape and the subject's response were recorded on a two-channel tape recorder (Ampex 354) at 7 1/2 i.p.s. Thus both the reaction time and the response were available for later analysis. Each subject responded to one complete list of 200 utterances. After the test, each subject was asked which pairs of words he felt he did well on and which pairs he felt he could not tell apart.

The tapes of each subject's performance were analyzed by computer. First, the voltages on each tape were digitized on a Radiation Inc. Analog Data Conversion System 152. The Ohio State University Instruction and Research Computer Center's IBM S/360 Mod 75 computer was used for further processing. The computer was programmed to determine changes in voltage. The transition from silence to voltage on the response channel was interpreted as the beginning of a response. The response was then categorized as either a sine wave or a square wave. The second channel containing voice was scanned to determine the transition from silence to voltage. This was construed as the beginning of a signal. The difference between the beginning of the signal and the beginning of the response was considered to be reaction time.[1]

---

[1] Measuring reaction time to speech stimuli, which exist in time, presents a problem not encountered with measuring reaction time to visual stimuli, namely at what point the subject can be said to begin to respond. The subject may begin to respond during the presentation of the word or after he has heard the entire word. On the other hand, reaction time can be measured either from the beginning or the end of the word. For this experiment, I have chosen to measure reaction time from the beginning of the word, in full awareness that either decision creates difficulties.

However, because of technical difficulties with the recordings, not all responses by every subject could be recovered.

Results

Identification: The over-all scores, given in Table 1, indicate that subjects do not seem to be able to identify the words correctly at significantly above chance levels. These results are presented

TABLE 1
PER CENT CORRECT IDENTIFICATIONS

Word Pair	Total		List A	List B	Phonetics Students	Phonetically Untrained Students
	per cent correct	range of scores in per cent				
1. wade/ weighed	50.4	20-70	46.3	55.2	48.9	51.3
2. hose/ hoes	51.1	30-73	46.3	56.8	52.8	50.1
3. bard/ barred	50.6	30-75	53.6	47.1	52.1	49.6
4. pact/ packed	50.4	33-67	52.2	48.6	54.2	48.1
5. lax/ lacks	45.1	20-70	47.2	42.8	42.6	47.1
6. baste/ based	49.6	25-70	46.4	53.5	43.2	53.7
7. adds/ adze	46.2	25-75	43.5	49.0	46.5	45.3
8. mist/ missed	45.5	25-65	48.9	42.5	47.8	43.9
9. laps/ lapse	55.4	30-75	55.2	55.6	51.8	58.8
10. guest/ guessed	49.0	20-85	48.7	49.1	46.8	50.6

graphically in Fig. 7. Furthermore, phonetics students do not seem to perform significantly differently from phonetically untrained subjects.

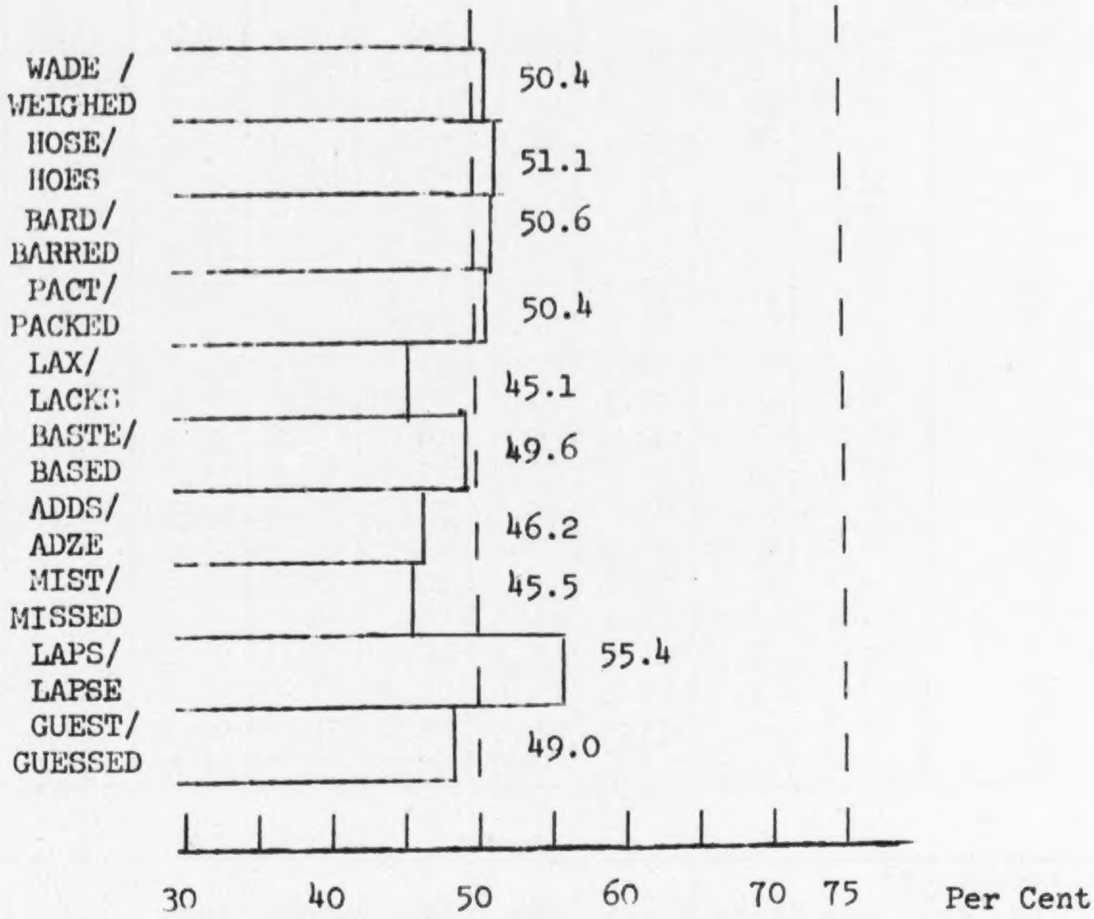


Fig. 7. Per Cent correct identifications for each word pair.

When the responses of the subjects to each production are analyzed, however, it appears that subjects are very consistent in their responses to some of the test items. Clearly consistent judgments (significant at .02 level or higher) for at least one production were obtained for the following pairs tested: weighed/wade, barred/bard, lax/lacks, baste/based, and mist/missed. Two pairs tested did not produce any significant agreement among subjects: hose/hoes and lapse/laps. Three pairs may or may not be considered significant; in each of these pairs, agreement in responses was reached for four productions at a .05 level of significance.

TABLE 2

CONSISTENCY OF SUBJECTS' RESPONSES
PER CENT OF S AGREEING IN RESPONSE B

(underlined scores are significant at .02 level)

List A

production number	wade	hose	bard	pact	lax	baste	adds	mist	lapse	guest
1	61.5	16.7	66.7	54.5	46.2	85.7	53.3	100.0	54.5	25.0
2	53.8	69.2	69.2	36.4	33.3	33.3	50.0	30.0	81.8	41.7
3	41.7	66.7	23.1	81.8	23.1	42.9	58.3	50.0	54.5	58.3
4	53.8	61.5	76.9	36.4	50.0	42.9	33.3	40.0	81.8	66.7
5	50.0	76.9	58.3	81.8	33.3	50.0	41.7	50.0	36.4	45.5
6	53.8	50.0	58.3	54.5	25.0	50.0	41.7	60.0	36.4	45.5
7	50.0	53.8	15.4	70.0	46.2	57.1	50.0	66.7	45.5	58.3
8	46.1	69.2	75.0	54.5	61.5	66.7	50.0	70.0	72.7	41.7
9	45.5	46.1	38.5	81.8	53.8	71.4	16.7	60.0	63.6	66.7
10	38.5	46.1	61.5	81.8	61.5	66.7	75.0	20.0	45.5	50.0
11	69.2	76.9	30.8	30.0	18.2	18.2	33.3	30.0	45.5	41.7
12	58.3	66.7	30.8	50.0	53.8	72.7	75.0	70.0	50.0	58.3
13	38.46	46.1	61.5	27.3	53.8	50.0	33.3	66.7	54.5	16.7
14	30.8	46.1	84.6	63.6	41.7	61.5	54.5	50.0	45.5	50.0
15	63.6	46.1	46.1	45.5	30.8	78.5	58.3	50.0	27.3	58.3
16	69.2	58.3	46.1	54.5	50.0	58.3	25.0	80.0	54.5	58.3
17	69.2	53.8	61.5	63.6	46.2	28.6	83.3	55.5	45.5	66.7
18	38.5	61.5	46.1	45.5	46.2	57.1	58.3	40.0	54.5	58.3
19	46.1	61.5	15.4	45.5	58.3	78.5	58.3	40.0	63.6	50.0
20	61.5	61.5	38.5	63.6	38.5	33.3	41.7	70.0	54.5	50.0

List B

production number	wade	hose	bard	pact	lax	baste	adds	mist	lapse	guest
1	63.6	45.5	27.3	60.0	58.3	16.7	75.0	63.6	71.4	42.9
2	54.5	55.5	45.5	20.0	50.0	36.4	50.0	45.5	28.6	69.2
3	63.6	55.6	54.5	55.6	50.0	58.3	25.0	63.6	50.0	61.5
4	54.5	50.5	18.2	70.0	54.5	33.3	25.0	63.6	50.0	50.0
5	60.0	27.3	66.7	50.0	45.6	75.0	75.0	54.5	71.4	57.1
6	54.5	60.0	60.0	55.6	63.6	50.0	16.7	36.4	57.1	21.4
7	36.4	36.4	30.0	50.0	41.7	66.7	50.0	50.0	42.9	64.3
8	54.5	81.8	10.0	44.4	45.6	50.0	41.7	50.0	61.5	69.2
9	54.5	10.0	50.0	77.8	83.3	20.0	83.3	45.5	42.9	28.6
10	100.0	45.5	45.5	60.0	41.7	54.5	50.0	72.7	64.3	35.7
11	66.7	55.5	63.6	60.0	54.5	25.0	58.3	70.0	50.0	28.6
12	50.0	70.0	36.4	60.0	36.4	50.0	36.3	27.3	50.0	21.4
13	36.4	45.5	63.6	25.0	66.7	41.7	54.5	27.3	30.8	42.9
14	63.6	36.4	30.0	50.0	33.3	75.0	58.3	81.8	46.2	21.4
15	60.0	55.6	18.2	30.0	33.3	45.5	50.0	27.3	71.4	61.5
16	63.6	55.6	54.5	50.0	41.7	33.3	54.5	27.3	28.6	53.8
17	54.5	63.6	30.0	57.1	66.7	66.7	45.5	45.5	64.3	64.3
18	18.2	40.0	36.4	70.0	22.2	25.0	50.0	45.5	57.1	42.9
19	72.7	27.3	40.0	55.6	63.6	72.7	63.6	63.6	42.9	57.1
20	72.7	60.0	45.5	20.0	66.7	50.0	25.0	63.6	35.7	58.3

The consistency of subjects' responses is represented in Table 2.

Even when subjects are highly consistent in agreeing on a particular response, they do not necessarily identify the word correctly; the identification scores for utterances for which subjects agree on one response (at .02 level) are still at chance level (57% correct).

Subject Interview: The mean identification score for the word pair judged easiest and for the most difficult word pair was calculated. The score represents each subject's performance in relation to his judgment of ease and difficulty, and thus does not represent performance on any one word pair. The differences found were not statistically significant, but did lie in an interesting direction: both phonetically trained and phonetically untrained subjects performed better on the word pairs they considered easy than on the word pairs they considered difficult.

TABLE 3

SUBJECTS' PERFORMANCE IN RELATION TO JUDGMENTS OF EASE AND DIFFICULTY

	Word Pair Judged Easiest (% Correct)	Word Pair Judged Most Difficult (% Correct)
All Subjects	53.10	46.01
Phonetics Students	51.60	49.20
Phonetically Untrained Students	54.10	43.80

Furthermore, subjects show a fair amount of agreement in judging which pairs of words are difficult and which are easy. Table 4 shows the number of times each word pair was judged easy and the number of times each word pair was judged difficult.

TABLE 4

EASE AND DIFFICULTY OF WORD PAIRS AS JUDGED BY SUBJECTS

Word pair	Number of times judged easy	Number of times judged difficult
wade/weighed	6	4
hose/hoes	3	7
bard/barred	5	1
pact/packed	1	2
lax/lacks	1	3
baste/based	3	2
adds/adze	1	5
mist/missed	3	0
laps/lapse	3	3
guest/guessed	3	1

Reaction time: Reaction time was not determined for all subjects. As Tables 5 to 8 show, reaction time was quite slow for all subjects and to all word pairs. There is no significant systematic difference in reaction time between correct and incorrect responses.

Reaction time to productions labeled consistently is quite variable. When the reaction time to consistently labeled productions is compared with the mean reaction time for that word pair, the differences in reaction time are in no way systematic. When the differences are statistically significant, however, then reaction time is longer to the consistently labeled production. These data are presented in Table 9.

When reaction time to mono-morphemic and to bi-morphemic words is examined, there is some tendency for reaction time to be shorter

TABLE 5

REACTION TIME, IN SECONDS, FOR SUBJECTS WITH TRAINING IN PHONETICS, FOR THE PAIRS WADE/WEIGHED, HOSE/HOES, BARD/BARRED, PACT/PACKED, AND LAX/LACKS

(mean and standard deviation; significantly different means for correct vs. incorrect responses are underlined)

Subject	wade / weighed		hose / hoes		bard / barred		pact / packed		lax / lacks	
D.G.										
all responses	1.180	.118	1.103	.250	.970	.152	1.229	.207	1.265	.222
correct responses	1.178	.129	1.170	.167	.916	.083	1.223	.257	1.289	.203
incorrect responses	1.183	.113	1.058	.292	1.024	.188	1.234	.169	1.245	.244
L.S.										
all responses	1.098	.199	1.181	.234	1.003	.180	--	--	1.192	.303
correct responses	.994	.153	1.187	.227	1.042	.168	--	--	1.147	.225
incorrect responses	1.159	.203	1.176	.252	.930	.194	--	--	1.220	.355
Z.B.										
all responses	--	--	--	--	--	--	--	--	--	--
correct responses	--	--	--	--	--	--	--	--	--	--
incorrect responses	--	--	--	--	--	--	--	--	--	--
S.Z.										
all responses	1.261	.398	1.814	.623	1.234	.546	1.651	.505	--	--
correct responses	1.348	.377	1.806	.649	1.320	.627	1.597	.518	--	--
incorrect responses	1.213	.419	1.822	.633	1.075	.348	1.698	.523	--	--

TABLE 6

REACTION TIME, IN SECONDS, FOR SUBJECTS WITH TRAINING IN PHONETICS, FOR THE PAIRS BASTE/BASED, ADDS/ADZE, MIST/MISSED, LAPS/LAPSE, AND GUEST/GUESSED

(mean and standard deviation; significantly different means for correct vs. incorrect responses are underlined)

Subject	baste / based		adds / adze		mist / missed		laps / lapse		guest / guessed	
D.G.										
all responses	1.096	.189	.957	.149	1.115	.227	1.137	.198	1.019	.081
correct responses	1.009	.063	.970	.199	1.158	.227	1.101	.158	.930	--
incorrect responses	1.143	.219	.948	.115	1.071	.178	1.190	.248	1.033	.078
L.S.										
all responses	--	--	--	--	.983	.159	1.369	.358	1.186	.294
correct responses	--	--	--	--	.973	.138	1.353	.417	1.188	.291
incorrect responses	--	--	--	--	.993	.189	1.402	.235	1.185	.324
Z.B.										
all responses	1.542	.440	1.468	.563	1.620	.525	1.589	.516	--	--
correct responses	1.465	.454	1.491	.586	1.603	.515	1.571	.406	--	--
incorrect responses	1.598	.443	1.442	.581	1.644	.599	1.597	.587	--	--
S.Z.										
all responses	1.263	.506	1.081	.330	--	--	1.243	.464	1.265	.509
correct responses	1.267	.292	1.049	.230	--	--	1.246	.413	1.088	.257
incorrect responses	1.261	.619	1.124	.445	--	--	1.241	.529	1.353	.590

TABLE 7

REACTION TIME, IN SECONDS, FOR PHONETICALLY UNTRAINED SUBJECTS FOR THE PAIRS WADE/WEIGHED, HOSE/HOES, BARD/BARRED, PACT/PACKED, AND LAX/LACKS

(mean and standard deviation; significantly different means for correct vs. incorrect responses are underlined)

Subject	wade / weighed		hose / hoes		bard / barred		pact / packed		lax / lacks	
1 all resp.	1.513	.281	1.377	.192	1.263	.241	1.376	.186	1.194	.327
correct resp.	1.403	.164	1.310	.176	1.214	.197	1.397	.199	1.180	.299
incorrect resp.	1.560	.312	1.421	.197	1.336	.295	1.350	.177	1.211	.375
2 all resp.	1.967	.516	2.458	.945	1.901	.642	2.382	.780	1.751	.618
correct resp.	1.725	.516	2.125	.502	2.181	.964	2.340	.718	1.631	.551
incorrect resp.	2.304	1.190	2.863	1.212	1.780	.439	2.431	.892	1.976	.714
3 all resp.	1.663	.816	2.150	.941	1.635	.515	1.687	.486	1.380	.569
correct resp.	1.443	.744	2.158	.941	1.570	.423	1.719	.521	2.045	.926
incorrect resp.	2.320	.734	--	--	1.690	.317	1.638	.473	1.158	.215
4 all resp.	2.056	.651	1.549	.667	1.323	.599	--	--	1.042	.639
correct resp.	2.061	.623	1.769	.768	1.227	.565	--	--	1.176	.793
incorrect resp.	2.050	.739	1.329	.502	1.498	.673	--	--	.922	.483
5 all resp.	1.734	.552	--	--	1.723	.460	2.139	.507	--	--
correct resp.	1.778	.532	--	--	1.741	.537	2.105	.459	--	--
incorrect resp.	1.640	.646	--	--	1.709	.425	2.162	.579	--	--
6 all resp.	--	--	1.867	.525	1.635	.268	--	--	--	--
correct resp.	--	--	1.778	.458	1.585	.191	--	--	--	--
incorrect resp.	--	--	1.993	.622	1.662	.309	--	--	--	--
7 all resp.	1.376	.228	1.778	.476	--	--	1.699	.272	1.628	.348
correct resp.	1.363	.168	1.779	.487	--	--	1.739	.290	1.443	.250
incorrect resp.	1.385	.278	1.776	.210	--	--	1.636	.258	1.814	.352
8 all resp.	1.972	.326	1.917	.336	--	--	2.364	.813	2.627	.664
correct resp.	1.910	.318	1.934	.329	--	--	2.778	.921	2.242	.305
incorrect resp.	2.057	.338	1.891	.376	--	--	2.123	.669	2.742	.709

TABLE 8

REACTION TIME, IN SECONDS, FOR PHONETICALLY UNTRAINED SUBJECTS. FOR THE PAIRS BASTE/BASED, ADDS/ADZE, MIST/MISSED, LAPS/LAPSE, AND GUEST/GUESSED

(mean and standard deviation; significantly different means for correct vs. incorrect responses are underlined)

Subject	baste / based		adds / adze		mist / missed		laps / lapse		guest / guessed	
1 all resp.	1.489	.267	--	--	--	--	--	--	--	--
correct resp.	1.431	.268	--	--	--	--	--	--	--	--
incorrect resp.	1.527	.271	--	--	--	--	--	--	--	--
2 all resp.	2.232	.543	--	--	--	--	--	--	--	--
correct resp.	2.305	.473	--	--	--	--	--	--	--	--
incorrect resp.	2.123	.652	--	--	--	--	--	---	---	--
3 all resp.	1.828	.859	1.280	.384	1.782	.729	1.725	.593	1.970	.476
correct resp.	2.163	.887	1.325	.247	1.793	.684	1.815	.604	1.984	.341
incorrect resp.	1.358	.617	1.267	.430	1.771	.828	1.235	.120	1.963	.548
4 all resp.	1.697	.754	1.523	.637	--	--	--	--	--	--
correct resp.	1.855	.735	1.400	.466	--	--	--	--	--	--
incorrect resp.	1.348	.749	1.625	.780	--	--	--	--	--	--
5 all resp.	--	--	--	--	--	--	--	--	1.845	.718
correct resp.	--	--	--	--	--	--	--	--	1.826	.566
incorrect resp.	--	--	--	--	--	--	--	--	1.857	.836
6 all resp.	--	--	.960	.185	.927	.134	1.103	.397	1.070	.296
correct resp.	--	--	.966	.154	.874	.149	.996	.264	1.136	.332
incorrect resp.	--	--	.954	.220	.951	.127	1.296	.547	.978	.241
7 all resp.	1.937	.385	1.742	.249	1.848	.373	1.652	.408	--	--
correct resp.	1.971	.441	1.748	.268	1.908	.434	1.506	.151	--	--
incorrect resp.	1.862	.246	1.729	.230	1.779	.305	1.972	.618	--	--
8 all resp.	2.073	.541	--	--	--	--	2.134	.311	2.135	.515
correct resp.	1.901	.385	--	--	--	--	2.283	.310	2.049	.468
incorrect resp.	2.345	.665	--	--	--	--	1.909	.129	2.292	.604

TABLE 9

REACTION TIME, IN SECONDS, TO PRODUCTIONS LABELED CONSISTENTLY
(reaction times significantly different from mean are underlined)

Subject	Mean RT for word pair	RT to production	Mean RT for word pair	RT to production			Mean RT to word pair	RT to production	Mean RT to word pair	RT to production	Mean RT to word pair	RT to production
		10 wade		7 bard	14 bard	19 barred		11 lax		1 based		1 mist
3	1.663	.720										
4	2.056	1.700										
5	1.734	1.681										
6	--	--										
7	1.376	1.527										
8	1.972	2.100										
D.G.			.970	.870	.970	.950	1.265	1.300	1.096	1.600	1.115	1.730
L.S.			1.003	1.194	.944	.994	--	--	--	--	.983	1.090
Z.B.			--	--	--	--	--	--	1.542	2.281	1.620	--
S.Z.			1.234	1.760	--	1.080	--	--	1.263	3.070	--	--
1			1.263	1.420	1.250	1.250	1.194	.950	1.489	1.160	--	--
2			1.901	1.070	1.950	3.650	1.751	3.370	2.232	2.870	--	--

to the bi-morphemic word, as Fry discovered. The differences, however, are not statistically significant. These data are presented in Tables 10 to 12.

Acoustic analysis: In order to discover the acoustic cues that subjects were employing to arrive at consistent labeling, spectrograms were made of all productions that were labeled consistently. Spectrograms were also made of some productions for each word pair that were labeled at random, and of the production that immediately preceded the consistently labeled production. Spectrograms were made on a Kay Electric Company Sonagraph.

It was found that subjects were employing two types of cues: slight differences in consonant quality and differences in vowel duration. For the word pairs *baste*/*based*, *mist*/*missed*, and *lax*/*lacks*, subjects were responding to a slight difference in the fricative [s]. The consistently labeled productions had more energy, at all frequencies, in the fricative than the productions that were labeled at random.

The word pairs *wade*/*weighed* and *bard*/*barred* were labeled consistently on the basis of vowel duration. However, subjects apparently were not responding to absolute differences in vowel duration, but to the duration of a vowel compared to the duration of the vowel of the preceding production. Thus a production [baɹd] would be labeled *barred* if it followed a production with a perceptibly shorter vowel; it would be labeled *bard* if it followed a production with a perceptibly longer vowel. It did not matter whether the word was intended as "bard" or "barred."

TABLE 10

REACTION TIME, IN SECONDS, TO THE MONO-MORPHEMIC AND BI-MORPHEMIC WORDS
WEIGHED/WADE, HOES/HOSE, BARRED/BARD. PACKED/PACT
(mean and standard deviation; significantly different means are underlined)

Subj.	weighed		wade		hoes		hose		barred		bard		packed		pact	
DG	1.129	.091	1.227	.152	1.083	.110	1.258	.181	.908	.080	.924	.095	1.206	.262	1.245	.289
LS	1.106	.094	.910	.138	1.152	.127	1.210	.285	1.039	.215	1.046	.115	--	--	--	--
ZB	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
SZ	1.255	.384	1.535	.403	1.765	.530	1.817	.718	1.152	.448	1.522	.799	1.830	.521	1.286	.383
1	1.320	.064	1.445	.191	1.377	.115	1.270	.205	1.179	.186	1.264	.223	1.305	.100	1.508	.241
2	1.652	.479	1.878	.626	2.278	.594	1.942	.333	2.990	.436	1.778	.778	2.473	.634	2.182	.854
3	1.464	.493	1.403	1.207	1.870	.398	2.590	1.612	1.608	.478	1.420	--	2.108	.604	1.498	.507
4	1.940	.747	2.142	.586	2.090	.439	1.576	.902	1.542	.718	.965	.211	--	--	--	--
5	1.798	.693	1.775	.317	1.745	.426	1.826	.567	1.860	.475	1.025	--	1.742	.233	2.468	.222
6	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
7	1.441	.113	1.312	.199	1.908	.574	1.606	.372	1.466	.010	1.764	.092	1.916	.315	1.632	.245
8	1.873	--	1.970	.329	2.074	.441	1.864	.278	--	--	--	--	2.299	.886	3.415	.546

TABLE 11
REACTION TIME, IN SECONDS, TO THE MONO-MORPHEMIC AND BI-MORPHEMIC WORDS LACKS/LAX. BASED/BASTE. ADDS/ADZE

(mean and standard deviation; significantly different means are underlined)

Subject	lacks		lax		based		baste		adds		adze	
DG	1.256	.269	1.330	.096	1.003	.056	1.017	.084	1.018	.134	.923	.261
LS	1.289	.069	.935	.205	--	--	--	--	--	--	--	--
ZB	--	--	--	--	1.679	.411	1.467	.408	1.459	.700	1.524	.556
SZ	--	--	--	--	1.255	.092	1.272	.354	1.104	.307	.980	.067
1	1.446	.213	.958	.011	1.555	.262	1.308	.239	--	--	--	--
2	1.453	.429	1.896	.676	2.190	.498	2.420	.461	--	--	--	--
3	2.045	.926	--	--	2.647	1.215	1.800	.422	1.500	--	1.150	--
4	1.042	.725	1.400	1.015	2.292	.525	1.332	.669	1.327	.418	1.510	.693
5	--	--	--	--	--	--	--	--	--	--	--	--
6	--	--	--	--	--	--	--	--	.862	.101	1.070	.126
7	1.435	.320	1.305	--	1.948	.449	1.997	.482	1.673	.148	1.879	.398
8	1.966	--	2.380	.269	2.175	.295	1.672	.296	--	--	--	--

TABLE 12

REACTION TIME, IN SECONDS, TO THE MONO-MORPHEMIC AND BI-MORPHEMIC WORDS
MISSED/MIST, LAPS/LAPSE, GUESSED/GUEST
(mean and standard deviation: significantly different means are underlined)

Subject	missed		mist		laps		lapse		guessed		guest	
DG	1.070	.150	1.246	.271	1.088	.143	1.115	.184	.930	--	--	--
LS	.927	.189	1.010	.102	1.110	.130	1.515	.473	1.188	.291	--	--
ZB	1.461	.515	1.660	.563	1.197	--	1.758	.346	--	--	--	--
SZ	--	--	--	--	.966	.168	1.665	.205	1.030	.226	1.127	.318
1 1	--	--	--	--	--	--	--	--	--	--	--	--
2	--	--	--	--	--	--	--	--	--	--	--	--
3	1.735	.007	1.816	.836	1.630	.241	1.968	.787	2.155	.191	1.870	.407
4	--	--	--	--	--	--	--	--	--	--	--	--
5	--	--	--	--	--	--	--	--	--	--	1.885	.677
6	1.010	.127	.783	.075	.910	.101	1.038	.084	1.260	.495	1.043	.173
7	2.295	--	1.859	.438	1.463	.106	1.531	.174	--	--	--	--
8	--	--	--	--	2.171	.339	2.339	.311	1.908	.368	2.219	.559

Discussion

To a great extent, the results of this experiment are negative. Subjects can not identify the word pairs correctly. They do not perform better on the word pairs they consider easy than on the word pairs they consider difficult. And no inferences can be drawn from the reaction time except that, because the reaction time is very slow, the subjects find it difficult to decide which word they have heard.

However, subjects seem to be aware of at least some sub-phonemic information since they label some word pairs consistently, even though not correctly. Faced with the task of the experiment, subjects develop a strategy for making use of fine phonetic detail. In this manner they arrive at some consistent labelings. But since the identifications based on this strategy are equally likely to be correct or incorrect, the strategy can not be considered to be part of ordinary speech perception.

Thus the results of the experiment imply that even though subjects may become aware of sub-phonemic differences, they do not know what linguistic use to make of them.

CHAPTER THREE

THE PERCEPTION OF OBSTRUENT CLUSTERS

Studies dealing with the perception of order of non-speech sounds indicate that perceiving the order of sounds of short duration is quite problematic. Hirsch (1959) reported that, after considerable practice, subjects could perceive the order of two sounds correctly if the onset of the sounds was separated by 15 to 20 msec. For stimuli, Hirsch used tones and bursts of noise 500 msec. in duration as well as clicks. Hirsch concludes that the minimal temporal interval required for perception of order is independent of the duration of the sound (within the limits of the experiment) and of the quality of the sound.

Broadbent and Ladefoged (1959) found that, at first, subjects could not perceive the order of sounds unless the onset of the sounds was separated by 150 msec.; with considerable training, a 30 msec. separation became adequate for accurate perception of order. Broadbent and Ladefoged used three different stimuli: a "hiss," high frequency noise of 120 msec. duration; a "pip," an 800 cps sine wave of 30 msec. duration, and a "buzz," a 171 cps square wave of 30 msec. duration.

Both these experiments involved the perception of the order of only two elements. However, the task is much more difficult when the subject has to determine the order of three or more elements. Several experiments involving the ordering of more than two sounds are reported by Warren and Warren (1970). In the first experiment, subjects were asked to determine the order of three sounds--a hiss, a tone, and a buzz, each lasting 200 msec.--which were repeated over and over without pauses. The subjects performed no better than chance. When the order of four sounds--a high tone, a low tone, a buzz, and a hiss, each lasting 200 msec.--was to be judged, the duration of each item had to be increased to between 300 and 700 msec. for half of the subjects to identify the sequence correctly. In the last experiment, the subjects were asked to judge the order of four 200 msec. vowel segments, cut from productions of extended vowels and spliced together without pauses. The subjects performed no better than chance. Identification of order became possible only when a 50 msec. silent interval was introduced between the vowels.

These experiments show that subjects have considerable difficulty in perceiving the order of sounds. However, listeners have no comparable difficulty with the order of elements in perceiving speech, even though many speech sounds are of quite short duration. Words like tax and task, ax and ask are normally perceived correctly, even though the duration of the consonants in the cluster is close to the minimum discovered in the Hirsch experiment. A reasonable estimate of the duration of p, t, and k is 51 msec., 30 msec. and 36 msec., respectively (Lehiste, 1970). These figures are derived from Estonian short voiceless stops.

It is, of course, a common observation that children have difficulty with such clusters; aks is a very common child pronunciation of ask, for example. And historically, such clusters have been prone to metathesis.[1] Still, adults seem to have no trouble with these clusters in the ordinary use of speech.

The observation that children have trouble with obstruent clusters but adults do not could imply that the adults' proficiency is a result of considerable practice. Both the Broadbent and Ladefoged, and Hirsch experiments show that the perception of order improves with practice. Analogously, the adults' proficiency could be a result of practice acquired in the course of language learning. However, it is also possible, and has been suggested by a number of theorists, that some special mechanisms are employed in the perception of consonant clusters. Thus Broadbent and Ladefoged report that the introspective feeling, developed in judging order, is that two items become differentiated on the basis of over-all quality rather than order. They suggest that the perceptual mechanism operates on discrete samples of perceptual information; when two items fall into the same sample their order has to be inferred on some other basis. On the basis of the Broadbent and Ladefoged and Hirsch experiments, Neisser (1967) argues that a listener gradually learns to distinguish a cluster like ts from a cluster like st, rather than perceiving a sequence of t followed by s, or s followed by t. He implies that such clusters are perceptual units to the listener, not normally analyzed further.

---

[1] It may be that the sporadic occurrence of metathesis, found in historical change, could be better explained by examining errors in perception rather than errors in production, which has been the traditional starting point for discussing language change.

Wickelgren's idea of context sensitive coding, presented in detail in Chapter One (Wickelgren, 1969a, 1969b), can also explain the fact that adults easily perceive a sequence of consonants correctly. When a listener is presented with a consonant cluster, e.g. sk, he knows that it is composed of two elements , but he does not encode these elements in order; rather, the cluster is coded as an unordered sequence, with each element identified for what precedes and follows it. Schematically, the coding would be something like the following: $s^{k}\#$ $\#^{s}k$. These elements can be assembled in the correct order, and the listener can arrive at the intended sequence.

The perception of obstruent clusters is an interesting problem for empirical study, particularly since it is related to the almost universally accepted notion that the minimal unit in speech perception is the phoneme. Both Neisser's suggestion and Wickelgren's theory, if substantiated, would argue against this view.

An experiment was designed to investigate the perceptual mechanisms employed in the perception of obstruent clusters. By observing the pattern of confusions of obstruent clusters in the presence of noise, it is possible to make some inferences about the perceptual mechanisms underlying the perception of these clusters.

Method

Stimuli: Fifteen pairs of English words were selected which differed from each other only in the order of obstruents in a cluster. Five pairs of words ended in the obstruent cluster ps/sp; five ended in ts/st; five ended in ks/sk. For each obstruent cluster, there was one pair of two-syllable words; in addition, each obstruent cluster appeared at least once with no morpheme boundary in the cluster. The full list of words is reproduced below:

apse
asp

lips
lisp

Capsian
Caspian

claps
clasp

raps
rasp

Blatz
blast

mats
mast

blitzer
blister

boots
boost

coats
coast

ax
ask

tax
task

axing
asking

Max
mask

bricks
brisk

Three lists were constructed. On each list, each word appeared two times in random order; the order was arrived at by using a table of random numbers. Thus each list consisted of 60 words; each consonant cluster appeared on each list ten times.

The speaker was a male, with a medium-pitch voice, from Akron, Ohio. Before recording, the speaker practiced for some time so that he could produce the stressed vowel of each word at a constant intensity. This was accomplished by monitoring the v.u. meter on the tape recorder. When the speaker was producing the words at a constant intensity, the actual recording was made, monitoring each production to keep the intensity at a constant level. The three lists were recorded in a sound-proof recording booth on an Ampex 350 tape recorder, at 7 1/2 i.p.s. Words were separated by 2.5 seconds; after every five words, there was a gap of 5 seconds.

The stimulus tape was made by re-recording the master tape while adding "white" noise produced by a Grayson-Stadler noise generator.

The instrumentation is shown in the accompanying diagram (Fig. 8).

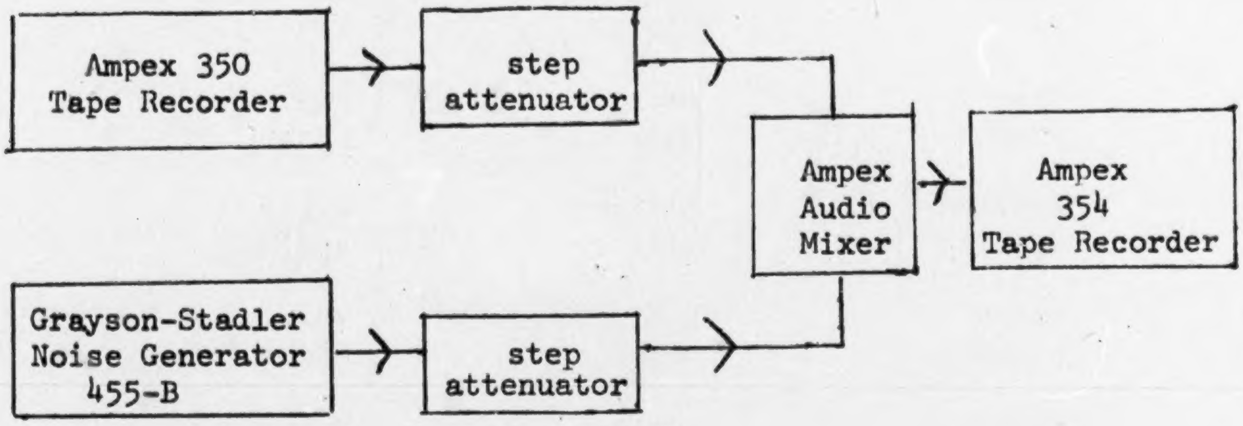

Ampex 350
Tape Recorder
step
attenuator
Ampex
Audio
Mixer
Ampex
354
Tape Recorder
Grayson-Stadler
Noise Generator
455-B
step
attenuator

Fig. 8. Instrumentation for adding noise to stimulus tape.

Three different signal-to-noise ratios were employed for the three lists: the first list was re-recorded at a signal-to-noise ratio of 0 d.b.; the second list was re-recorded at a signal-to-noise ratio of +12 d.b.; the third list was recorded at a signal-to-noise ratio of -6 d.b.

Subjects: Nineteen subjects participated in the experiment. All were members of The Ohio State University linguistics department and native speakers of English.

Procedure: The experiment was conducted as a listening test. Before the test, subjects were instructed to write what they heard, and to guess if necessary; they were told to expect some unusual words, and these words were shown to them. For the test, the stimulus tape was played on a tape recorder while the subjects listened over earphones, and wrote what they heard on an answer sheet. Each subject listened to the entire tape (3 lists), and thus responded to 180 stimulus words.

In addition, five subjects took the test a second time. In the second test, the listening conditions were identical to those of the first test, but the subjects were instructed to say what they heard. The subjects' spoken response and the stimulus tape were recorded on separate channels of an Ampex 354 tape recorder.

The subjects' responses were tabulated in the form of confusion matrices. The answers were scored only for the perception of the obstruent clusters. Thus, if the stimulus word was raps, but the subject wrote laps, he was scored correct.

The response tapes of the five subjects who gave spoken responses were processed by an Elema-Schönander Mingograf, each channel of the tape being represented as an oscillogram on a separate channel of the Mingograf. The paper speed was 100 mm/sec.

Reaction time was determined by measuring from the onset of the stimulus word to the onset of the response, and from the end of the stimulus word to the onset of the response. There was no difficulty in measurement when the signal-to-noise ratio was +12 d.b. When the signal-to-noise ratio was 0 d.b., measurements from the stimulus word had to be made from the vowel rather than from the consonants. Reaction time could not be determined when the signal-to-noise ratio was -6 d.b.

Results

Confusions: The results are presented in the accompanying confusion matrices (Tables 13 to 51). Each cell of the matrices shows the number of times the stimulus consonant cluster, given at the beginning of the row, was identified as the consonant cluster given in the column heading. Correct responses lie on the diagonal. In addition, the percent of all the responses of each row that lie in a particular cell is given for each cell. A.I. (articulation index) gives the ratio of correct identifications for each matrix.

Tables 13 to 15 give confusion matrices for all responses. As is to be expected, the higher the noise is, in relation to the signal, the more confusion errors occur. It can be observed that, for all consonant clusters, the most common error is a reversal of the consonant cluster. Furthermore, the stop-fricative cluster is perceived correctly more often than the corresponding fricative-stop cluster. This effect may result from the higher frequency of stop-fricative clusters in English.

The pattern of confusions for written responses (Tables 16 to 18) and for spoken responses (Tables 19 to 21) is essentially the same. Thus, there is no advantage to spoken responses, and spoken responses do not produce a different pattern of confusions.

Tables 22 to 27 present the confusion matrices for two-syllable words. The articulation index is slightly higher for two-syllable words, but the confusion patterns remain essentially the same. There is some tendency to confuse p̲ and k̲ clusters only with each other, and not with t̲ clusters; however, this is probably due to other differences in the two-syllable words tested, i.e., a different vowel and a different final consonant.

Tables 28 to 45 present confusion matrices for all test words with a given vowel. The most common confusion, for all vowels, is still a reversal of the consonant cluster. There is only one exception to this tendency; when the vowel is [ɪ], p̲ clusters tend to be confused with t̲ clusters about as much as with each other.

TABLE 13
ALL RESPONSES--SIGNAL TO NOISE RATIO: +12 d.b.

AI: .8599

	TS	ST	PS	SP	KS	SK
TS	81.9 181	9 20	2.3 5		1.4 3	5.4 12
ST	6.5 15	78.8 182		1.3 3	.9 2	12.5 29
PS	1.7 4	.8 2	86.2 206	8.4 20	2.5 6	.4 1
SP	.4 1	3.5 8	3.5 8	77.3 177		15.3 35
KS		.4 1	1.3 3	.9 2	95.2 219	2.2 5
SK	1.3 3		1.3 3	.8 2	.4 1	96.2 226

TABLE 14
ALL RESPONSES--SIGNAL TO NOISE RATIO: 0 d.b.

AI: .4896

	TS	ST	PS	SP	KS	SK
TS	54.3 109	33.3 67	2.9 6	.5 1	.9 1	7.9 16
ST	45.3 87	39.6 76	5.2 10	.5 1	.5 1	8.9 17
PS	10.2 19	7 13	44.6 83	27.4 51	3.3 6	7.5 14
SP	14.5 27	7 13	23.7 44	37.1 69	4.3 8	13.4 25
KS	9.1 16	3.4 6	1.7 3	2.8 5	67.6 119	15.4 27
SK	7.7 13	2.9 5	14.3 24	5.4 9	17.9 30	51.8 87

TABLE 15

ALL RESPONSES--SIGNAL TO NOISE RATIO: -6 d.b.

AI: .3874

	TS	ST	PS	SP	KS	SK
TS	57.2 88	29.2 45	1.9 3	1.3 2	5.8 9	4.6 7
ST	55.7 78	32.2 45	1.4 2	1.4 2	5 7	4.3 6
PS	10.4 18	4.6 8	36.4 63	32.5 46	9.2 16	6.9 12
SP	19.7 29	13.6 20	13.6 20	29.9 44	11.6 17	11.6 17
KS	8.5 12	3.5 5	9.2 13	7.1 10	38.4 54	33.3 47
SK	4.9 5	2.9 3	14.7 15	15.7 16	24.5 25	37.3 38

TABLE 16

ALL WRITTEN RESPONSES--SIGNAL TO NOISE RATIO: +12 d.b.

AI: .8529

	TS	ST	PS	SP	KS	SK
TS	81.4 139	10.5 18	2.3 4		1.2 2	4.6 8
ST	7.7 14	77.9 141		1.7 3	1.1 2	11.6 21
PS	1.6 3	1.1 2	85 160	9.6 18	2.1 4	.6 1
SP	.6 1	2.8 5	3.9 7	76.9 137		15.8 28
KS		.6 1	1.6 3	1.1 2	94.5 169	2.2 4
SK	1.7 3		1.7 3	.5 1	.5 1	95.6 176

TABLE 17

ALL WRITTEN RESPONSES--SIGNAL TO NOISE RATIO: 0 d.b.

AI: .5006

	TS	ST	PS	SP	KS	SK
TS	55.5 87	32.5 51	3.1 5	.6 1	1.3 2	7 11
ST	44.5 65	41.1 60	5.5 8	.7 1	.7 1	7.5 11
PS	7.9 11	7.9 17	42.5 59	29.5 41	2.9 4	9.3 13
SP	11.4 16	9.2 13	24.8 35	36.9 52	4.9 7	12.8 18
KS	8.9 12	3.7 5	2.2 3	1.5 2	68.6 92	14.9 20
SK	5.6 7	3.1 4	11.1 14	5.6 7	17.5 22	57.1 72

TABLE 18
ALL WRITTEN RESPONSES--SIGNAL TO NOISE RATIO: -6 d.b.

AI: .4121

	TS	ST	PS	SP	KS	SK
TS	56.7 63	36 34	1.8 2	1.8 2	7.2 8	1.8 2
ST	51 52	40.2 41	.98 1	1.96 2	2.94 3	2.94 3
PS	8.86 11	5.64 7	37.9 47	31.4 39	9.7 12	6.5 8
SP	15.3 16	13.3 14	16.2 17	34.3 36	12.4 13	8.6 9
KS	8.2 8	4.1 4	11.2 11	7.2 7	37.8 37	31.6 31
SK	5.8 4	4.4 3	14.5 10	15.9 11	20.3 14	39.2 27

TABLE 19
TOTAL SPOKEN RESPONSES--SIGNAL TO NOISE RATIO: +12 d.b.

AI: .8849

	TS	ST	PS	SP	KS	SK
TS	84 42	4 2	2 1		2 1	8 4
ST	2 1	82 41				16 8
PS	1.9 1		90.4 46	3.9 2	3.9 2	
SP		5.9 3	1.9 1	78.5 40		13.7 7
KS					98.2 50	1.9 1
SK				1.9 1		98.2 50

TABLE 20
TOTAL SPOKEN RESPONSES--SIGNAL TO NOISE RATIO: 0 d.b.

AI: .4548

	TS	ST	PS	SP	KS	SK
TS	50 22	36.4 16	2.3 1			11.3 5
ST	47.9 22	34.8 16	4.35 2			13.02 6
PS	17 8	4.25 2	51 24	21.3 10	4.25 2	2.13 1
SP	24.4 11		20 9	37.8 17	2.2 1	15.6 7
KS	9.6 4	2.4 1		6.7 3	64.4 27	16.7 7
SK	14.3 6	2.4 1	23.8 10	4.8 2	19 8	35.7 15

TABLE 21

TOTAL SPOKEN RESPONSES--SIGNAL TO NOISE RATIO: -6 d.b.

AI: .3266

	TS	ST	PS	SP	KS	SK
TS	58.2 25	25.6 11	2.3 1		2.3 1	11.6 5
ST	68.5 26	10.5 4	2.6 1		10.5 4	7.9 3
PS	14.3 7	20.4 1	32.7 16	34.7 17	8.2 4	8.2 4
SP	31 13	14.3 6	6.7 3	19 8	9.6 4	19 8
KS	9.3 4	2.3 1	4.7 2	6.9 3	39.6 17	37.2 16
SK	3 1		15.2 5	15.2 5	33.3 11	33.3 11

TABLE 22
WRITTEN RESPONSES FOR TWO-SYLLABLE WORDS
SIGNAL TO NOISE RATIO: +12 d.b.
(blister/blitzer, Capsian/Caspian, axing/asking)
AI: .9598

	TS	ST	PS	SP	KS	SK
TS	88.9 32	8.3 3			2.8 1	
ST	2.7 1	97.3 36				
PS			97.5 39	2.5 1		
SP		2.6 1		97.4 37		
KS			2.9 1		97.1 34	
SK				2.6 1		97.4 37

TABLE 23
WRITTEN RESPONSES FOR TWO-SYLLABLE WORDS
SIGNAL TO NOISE RATIO: 0 d.b.
(blister/blitzer, Capsian/Caspian, axing/asking)
AI: .5730

	TS	ST	PS	SP	KS	SK
TS	40.7 11	51.9 14				7.4 2
ST	30.6 11	66.7 24	2.7 1			
PS			37.9 11	58.7 17		3.4 1
SP			25.8 8	67.7 21		6.5 2
KS	4 1			8 2	64 16	24 6
SK	3.3 1		6.7 2	16.7 5	10 3	63.3 19

TABLE 24
SPOKEN RESPONSES FOR TWO-SYLLABLE WORDS
SIGNAL TO NOISE RATIO: +12 d.b.
(blister/blitzer, Capsian/Caspian, axing/asking)
AI: .95

	TS	ST	PS	SP	KS	SK
TS	100 10					
ST		100 10				
PS			100 10			
SP			10 1	90 9		
KS					90 9	10 1
SK				10 1		90 9

TABLE 25
SPOKEN RESPONSES FOR TWO-SYLLABLE WORDS
SIGNAL TO NOISE RATIO: 0 d.b.
(blister/blitzer, Capsian/Caspian, axing/asking)
AI: .636

	TS	ST	PS	SP	KS	SK
TS	50 5	50 5				
ST	50 5	50 5				
PS			70 7	30 3		
SP			10 1	70 7	20 2	
KS				16.7 1	83.3 5	
SK			11.1 1	11.1 1	11.1 1	66.7 6

TABLE 26
WRITTEN RESPONSES FOR TWO-SYLLABLE WORDS
SIGNAL TO NOISE RATIO: -6 d.b.
(blister/blitzer, Capsian/Caspian, asking/axing)
AI: .4524

	TS	ST	PS	SP	KS	SK
TS	50 7	50 7				
ST	50 10	45 9	5 1			
PS			40.6 13	50 16	3.1 1	6.3 2
SP			19.1 4	71.4 15		9.5 2
KS	4.8 1		28.5 6	23.8 5	23.8 5	19.1 4
SK				44.4 8	11.1 2	44.4 8

TABLE 27
SPOKEN RESPONSES FOR TWO-SYLLABLE WORDS
SIGNAL TO NOISE RATIO: -6 d.b.
(blister/blitzer, Capsian/Caspian, axing/asking)
AI: .3953

	TS	ST	PS	SP	KS	SK
TS	80 4	20 1				
ST	100 5					
PS			44.4 4	44.4 4		11.1 1
SP	10 1	20 2	10 1	50 5		10 1
KS				20 1	20 1	60 3
SK			11.1 1	33.3 3	22.2 2	33.3 3

TABLE 28
SPOKEN RESPONSES FOR [ɪ]--SIGNAL TO NOISE RATIO: +12 d.b.
(blister/blitzer, lips/lisp, brisk/bricks)
AI: .9000

	TS	ST	PS	SP	KS	SK
TS	100 10					
ST		100 10				
PS	10 1		70 7		20 2	
SP		20 2		70 7		10 1
KS					100 10	
SK						100 10

TABLE 29
SPOKEN RESPONSES FOR [æ]--SIGNAL TO NOISE RATIO: +12 d.b.
(mats/mast, Blatz/blast, ax/ask, apse/asp, Max/mask, tax/task, raps/rasp, claps/clasp, Capsian/Caspian, asking/axing)
AI: .8683

	TS	ST	PS	SP	KS	SK
TS	70 14	5 1	5 1			20 4
ST	4.8 1	57.1 12				38.1 8
PS			95.1 39	4.9 2		
SP		2.4 1	2.4 1	80.6 33		14.6 6
KS					97.6 40	2.4 1
SK				2.4 1		97.6 40

TABLE 30
SPOKEN RESPONSES FOR [u] AND [oʊ]--SIGNAL TO NOISE RATIO: +12 d.b.
(coats/coast, boots/boost)
AI: .9487

	TS	ST	PS	SP	KS	SK
TS	90 18	5 1			5 1	
ST		100 19				

TABLE 31
SPOKEN RESPONSES FOR [ɪ]--SIGNAL TO NOISE RATIO: 0 d.b.
(blister/blitzer, lips/lisp, bricks/brisk)
AI: .4655

	TS	ST	PS	SP	KS	SK
TS	50 5	50 5				
ST	50 5	50 5				
PS	30 3	10 1	20 2	20 2	20 2	
SP	40 4		20 2	40 4		
KS					90 9	10 1
SK			25 2		50 4	25 2

TABLE 32
SPOKEN RESPONSES FOR [æ]--SIGNAL TO NOISE RATIO: 0 d.b.
(matz/mast, Blatz/blast, ax/ask, apse/asp, Max/mask, tax/task, raps/rasp, claps/clasp, Capsian/Caspian, asking/axing)
AI: .4412

	TS	ST	PS	SP	KS	SK
TS	28.6 4	28.6 4	7.1 1			35.7 5
ST	27.8 5	27.8 5	11.1 2			33.3 6
PS	13.5 5	2.7 1	59.5 22	21.6 8		2.7 1
SP	20 7		20 7	37.1 13	2.9 1	20 7
KS	12.5 4	3.1 1		9.4 3	56.3 18	18.7 6
SK	17.6 6	2.8 1	23.6 8	5.4 2	11.8 4	38.2 13

TABLE 33
SPOKEN RESPONSES FOR [u] AND [oʊ]--SIGNAL TO NOISE RATIO: 0 d.b.
(coast/coats, boost/boats)
AI: .5000

	TS	ST	PS	SP	KS	SK
TS	65 13	35 7				
ST	66.7 12	33.3 6				

TABLE 34
SPOKEN RESPONSES FOR [ɪ]--SIGNAL TO NOISE RATIO: -6 d.b.
(blister/blitzer, lips/lisp, bricks/brisk)
AI: .2791

	TS	ST	PS	SP	KS	SK
TS	80 4	20 1				
ST	83.3 5				16.7 1	
PS	40 4	10 1	10 1	10 1	20 2	10 1
SP	66.7 6	11 1	11.1 1	11.1 1		
KS		11.1 1			44.4 4	44.4 4
SK			33.3 1		33.3 1	66.7 2

TABLE 35
SPOKEN RESPONSES FOR [æ]--SIGNAL TO NOISE RATIO: -6 d.b.
(mats/mast, Blatz/blast, ax/ask, apse/asp, Max/mask, tax/task, raps/rasp, claps/clasp, Capsian/Caspian, asking/axing)
AI: .3136

	TS	ST	PS	SP	KS	SK
TS	45 9	20 4	5 1		5 1	25 5
ST	50 7		7.2 1		21.4 3	21.4 3
PS	7.7 3		38.5 15	41 16	5.1 2	7.7 3
SP	21.2 7	15.2 5	6.1 2	21.2 7	12.1 4	24.2 8
KS	11.8 4		5.9 2	8.8 3	38.2 13	35.3 12
SK	3.5 1		13.8 4	17.2 5	34.5 10	31 9

TABLE 36
SPOKEN RESPONSES FOR [u] AND [oʊ]--SIGNAL TO NOISE RATIO: -6 d.b.
(coats/coast, boots/boost)
AI: .4444

	TS	ST	PS	SP	KS	SK
TS	66.7 12	33.3 6				
ST	77.8 14	22.2 4				

TABLE 37
WRITTEN RESPONSES FOR [æ]--SIGNAL TO NOISE RATIO: -6 d.b.
(mats/mast, Blatz/blast, ax/ask, apse/asp, Max/mask, tax/task, raps/rasp, claps/clasp, Capsian/Caspian, asking/axing)

AI: .4117

	TS	ST	PS	SP	KS	SK
TS	55.3 21	13.2 5	2.6 1	2.6 1	21 8	5.3 2
ST	34.6 9	30.9 8	3.8 1	7.7 2	11.5 3	11.5 3
PS	1.1 1	3.3 3	42.5 39	35.8 33	11.9 11	5.4 5
SP	6.7 5	9.3 7	18.7 14	40 30	16 12	9.3 7
KS	10.3 7	4.4 3	13.2 9	8.8 6	39.8 27	23.5 16
SK	3.4 2	5.2 3	15.5 9	19 11	19 11	37.9 22

TABLE 38
WRITTEN RESPONSES FOR [ɪ]--SIGNAL TO NOISE RATIO: -6 d.b.
(lips/lisp, bricks/brisk, blister/blitzer)

AI: .3094

	TS	ST	PS	SP	KS	SK
TS	33.3 6	55.6 10	5.6 1	5.6 1		
ST	55.6 10	44.4 8				
PS	31.3 10	12.5 4	25 8	18.7 6	3.1 1	9.4 3
SP	36.6 11	23.3 7	10 3	20 6	3.3 1	6.7 2
KS	3.3 1	3.3 1	6.7 2	3.3 1	33.3 10	50 15
SK	18.2 2		9.1 1		27.3 3	45.4 5

TABLE 39
WRITTEN RESPONSES FOR [u] AND [oʊ]--SIGNAL TO NOISE RATIO: -6 d.b.
(boots/boost, coats/coast)

AI: .5398

	TS	ST	PS	SP	KS	SK
TS	65.5 36	44.5 19				
ST	56.9 33	53.1 25				

TABLE 40
WRITTEN RESPONSES FOR [ɪ]--SIGNAL TO NOISE RATIO: 0 d.b.
(lips/lisp, bricks/brisk, blister/ blitzer)
AI: .5515

	TS	ST	PS	SP	KS	SK
TS	50 16	43.8 14				6.2 2
ST	22.8 8	77.2 27				
PS	19.4 6	16.1 5	32.3 10	29 9		3.2 1
SP	17.8 5	17.8 5	21.5 6	42.9 12		
KS	5.3 1				94.7 18	
SK	5 1		5 1	5 1	45 9	40 8

TABLE 41
WRITTEN RESPONSES FOR [æ]--SIGNAL TO NOISE RATIO: 0 d.b.
(mats/mast, Blatz/blast, ax/ask, apse/asp, Max/mask, tax/task, raps/rasp, claps/clasp, Capsian/Caspian, asking/axing)
AI: .4991

	TS	ST	PS	SP	KS	SK
TS	59.3 45	18.4 14	6.6 5	1.3 1	2.6 2	11.8 9
ST	40.3 23	26.3 15	12.3 7	1.8 1		19.3 11
PS	4.6 5	5.6 6	45.4 49	29.7 32	3.6 4	11.1 12
SP	9.7 11	7.1 8	25.6 29	35.4 40	6.2 7	16 18
KS	9.6 11	4.4 5	2.6 3	1.7 2	64.3 74	17.4 20
SK	5.6 6	3.8 4	12.3 13	5.6 6	12.3 13	60.4 64

TABLE 42
WRITTEN RESPONSES FOR [u] AND [ou]--SIGNAL TO NOISE RATIO: 0 d.b.
(boost/boots, coast/coats)
AI: .4271

	TS	ST	PS	SP	KS	SK
TS	53 26	47 23				
ST	63 34	33.3 18	1.9 1		1.9 1	

TABLE 43

WRITTEN RESPONSES FOR [æ]--SIGNAL TO NOISE RATIO: +12 d.b.

(mats/mast, Blatz/blast, ax/ask, apse/asp, Max/mask, tax/task, raps/rasp, claps/clasp, Capsian/Caspian, asking/axing)

AI: .8173

	TS	ST	PS	SP	KS	SK
TS	62.8 27	6.9 3	9.3 4		2.4 1	18.6 8
ST	17.9 12	43.3 29		4.5 3	2.9 2	31.4 21
PS	.6 1	1.3 2	86.5 133	11 17		.6 1
SP	.7 1	2.1 3	4.2 6	75.4 107		17.6 25
KS			2.1 3	1.4 2	93.7 134	2.8 4
SK	2.1 3		2.1 3	.7 1	.7 1	94.4 138

TABLE 44

WRITTEN RESPONSES FOR [ɪ]--SIGNAL TO NOISE RATIO: +12 d.b.

(lips/lisp, bricks/brisk, blister/blitzer)

AI: .8909

	TS	ST	PS	SP	KS	SK
TS	78.6 33	19 8			2.4 1	
ST	2.9 1	97.1 33				
PS	5.8 2		79.4 27	2.9 1	11.9 4	
SP		5.6 2	2.8 1	83.3 30		8.3 3
KS		2.8 1			97.2 35	
SK						100 38

TABLE 45

WRITTEN RESPONSES FOR [u] AND [ou]--SIGNAL TO NOISE RATIO: +12 d.b.

(boost/boots, coast/coats)

AI: .9518

	TS	ST	PS	SP	KS	SK
TS	91.9 79	8.1 7				
ST	1.2 1	98.8 79				

For both p and t, the second formant transition would be negative before [ɪ] (as opposed to k). Perhaps this fact accounts for the confusion.

Tables 46 to 51 present confusions for bi-morphemic words. Apparently, the presence of a morpheme boundary does not deter confusions; rather, mono-morphemic and bi-morphemic words produce similar confusion patterns.

Reaction time: Reaction time was compared for the two different signal-to-noise conditions, for words ending in different consonant clusters, and for correct vs. incorrect responses.

Reaction time was significantly faster when the signal-to-noise ratio was +12 d.b., than when the signal-to-noise ratio was 0 d.b.

As can be seen in Table 52, reaction time was consistently faster for correct responses than for incorrect responses, although the difference did not always reach statistical significance.

When the reaction time to the individual consonant clusters is examined, the reaction time is significantly slower to words ending in ps, sp, and sk clusters when the signal-to-noise ratio is 0 d.b. When the signal-to-noise ratio is +12 d.b., reaction time is significantly slower only to words ending in ps clusters.[2] (Table 53).

---

[2] This difference may be a result of the frequency of the words. For example, apse is not even listed in An English Word Count (Wright, 1965).

Finally, the reaction time to two-syllable words, when measured from the beginning of the word, is about the same as the reaction time to one-syllable words. When measured from the end of the word, the

TABLE 46

WRITTEN RESPONSES FOR BI-MORPHEMIC WORDS--SIGNAL TO NOISE RATIO: +12 d.b.

(lips, claps, naps, bricks, coats, mats, boots)
AI: .8391

	TS	ST	PS	SP	KS	SK
TS	79.3 73	9.8 9	4.3 4			6.6 6
PS	1.9 2		83.7 87	10.6 11	3.8 4	
KS		2.9 1			97.1 33	

TABLE 47

WRITTEN RESPONSES FOR BI-MORPHEMIC WORDS--SIGNAL TO NOISE RATIO: 0 d.b.

(lips, claps, naps, bricks, coats, mats, boots)
AI: .4798

	TS	ST	PS	SP	KS	SK
TS	45 36	35 28	5 4		1.3 1	13.7 11
PS	13.5 10	14.9 11	39.2 29	28.4 21	1.4 1	2.6 2
KS	5.3 1				94.7 18	

TABLE 48

WRITTEN RESPONSES FOR BI-MORPHEMIC WORDS--SIGNAL TO NOISE RATIO: -6 d.b.

(lips, claps, naps, bricks, coats, mats, boots)
AI: .4702

	TS	ST	PS	SP	KS	SK
TS	63.9 46	25 18	2.8 2			8.3 6
PS	20.9 17	6.2 5	34.6 28	29.7 24	4.9 4	3.7 3
KS		3.1 1	6.2 2	3.1 1	40.7 13	46.9 15

TABLE 49

SPOKEN RESPONSES FOR BI-MORPHEMIC WORDS--SIGNAL TO NOISE RATIO: +12 d.b.
(lips, claps, raps, bricks, coats, mats, boots)
AI: .8116

	TS	ST	PS	SP	KS	SK
TS	72.5 21	6.9 2	3.4 1		3.4 1	13.8 4
PS	3.3 1		83.3 25	6.7 2	6.7 2	
KS					100 10	

TABLE 50

SPOKEN RESPONSES FOR BI-MORPHEMIC WORDS--SIGNAL TO NOISE RATIO: 0 d.b.
(lips, claps, raps, bricks, coats, mats, boots)
AI: .4769

	TS	ST	PS	SP	KS	SK
TS	48.4 14	31 9			3.4 1	17.2 5
PS	26.9 7	7.7 2	30.8 8	26.9 7	7.7 2	
KS					90 9	10 1

TABLE 51

SPOKEN RESPONSES FOR BI-MORPHEMIC WORDS--SIGNAL TO NOISE RATIO: -6 d.b.
(lips, claps, raps, bricks, coats, mats, boots)
AI: .4194

	TS	ST	PS	SP	KS	SK
TS	50 13	26.9 7	3.8 1		3.8 1	15.5 4
PS	26.9 7		30.8 8	26.9 7	11.6 3	3.8 1
KS	10 1				50 5	40 4

TABLE 52

REACTION TIME, IN MSEC., FOR CORRECT AND INCORRECT RESPONSES

Consonant Cluster	Signal to Noise Ratio: 0 d.b.				Signal to Noise Ratio: +12 d.b.			
	Correct		Incorrect		Correct		Incorrect	
	Beginning	End	Beginning	End	Beginning	End	Beginning	End
1 syllable words:								
TS	1011	744	989	779	887	646	1012	739
ST	1209	939	957	734	863	619	946	664
2 syllable words:								
TS	954	519	1198	827	862	377	-	-
ST	1169	809	1102	746	853	376	-	-
1 syllable words:								
PS	1038	845	1167	980	960	747	1247	1063
SP	1042	839	1211	986	918	673	809	582
2 syllable words:								
PS	984	474	1025	514	875	343	-	-
SP	1095	633	919	412	876	312	930	360
1 syllable words:								
KS	1067	866	1074	862	837	619	-	-
SK	931	692	1194	945	918	679	-	-
2 syllable words:								
KS	1156	690	1018	655	826	319	1610	1190
SK	1008	489	1420	923	808	287	1175	695

TABLE 53
REACTION TIME, IN MSEC., TO CONSONANT CLUSTER
(mono-syllabic words only)

	Signal to Noise Ratio: 0 d.b.		Signal to Noise Ratio: +12 d.b.	
	Beginning	End	Beginning	End
TS	998	773	913	667
ST	1035	771	904	629
PS	1109	914	981	772
SP	1092	878	891	649
KS	1037	835	837	620
SK	1138	890	918	679

reaction time is much shorter to two-syllable words. Apparently, subjects begin to respond to the two-syllable words before they hear the whole word, probably as soon as they hear the medial consonant cluster.

Discussion

The finding that has the most bearing on the perception of consonant clusters is that reversal errors are the most common errors. This finding is counter to the idea that the phoneme is the minimal perceptual unit; if consonant clusters are perceived "phoneme-by-phoneme," then, when a listener hears the consonant cluster sp, he first hears s and then he hears p. Given that he hears these in a particular order, there is no reason for him to reverse that order. Granted, he might on occasion forget the order, but there is no reason to suppose that he would be more likely to forget the order of the consonants than to forget one of the consonants; thus, reversal errors would be no more common than substitution errors. However, that is clearly not the case: reversal errors are much more common. This finding implies that some special perceptual mechanisms must be postulated for the perception of consonant clusters.

Broadbent and Ladefoged's suggestion appears of doubtful validity, not because the consonant cluster data contradict it, but for other reasons. As has already been pointed out by Neisser, a listener is not limited to an invariant time-determined chunk of input that he can process. This is implied by the ability of listeners to perceive correctly speech that is speeded up. Broadbent and Ladefoged would have to claim that order errors would become more common, and involve more segments, as speech is speeded up, since each "time chunk" would contain more segments. But that this is not the case seems clear from personal experience with record players.

Neisser's suggestion, that a consonant cluster is a perceptual unit, and Wickelgren's suggestion that a consonant cluster is coded in terms of some element very much like an allophone, are both compatible with the data.

If consonant clusters are perceptual units, then clearly a *ps* cluster is most similar to a *sp* cluster. If this is so, then, when the signal is degraded by the addition of noise, the items that are most similar to each other will be confused most; thus, reversal errors will be most likely.

If a consonant cluster is coded in terms of allophones, then the allophone of *s* before *p* will be slightly different, acoustically, from the allophone of *s* after *p*. This difference, however, will be the most subtle part of the signal; particularly, it will be smaller than the acoustic information differentiating consonants from each other. These small acoustic differences will be the first to disappear when the signal is degraded by noise; consequently, reversal errors will be the most common in a degraded signal.

Thus, either Neisser's or Wickelgren's suggestion will account for the observed result.

CHAPTER FOUR

SYNTACTIC UNITS IN PERCEPTION

Experiments involving the localization of "clicks" in sentences have been used by Bever, Fodor, and others (Fodor and Bever, 1965; Bever, Lackner and Kirk, 1969) to examine syntactic units in perception. The experiments are based on a phenomenon discovered by Ladefoged and Broadbent (1960) that subjects have great difficulty localizing a click in speech, when the click and speech are presented simultaneously.

At first, the "click" experiments seemed to support the view that syntactic constituents were perceptual units: when asked to locate a click, subjects tended to move it towards a constituent boundary. A theory of perception was developed to explain the phenomenon: a subject could pay attention to one thing at a time, he could either process speech or the click; subjects would not interrupt perceptual units of speech; consequently, subjects would tend to locate the click between perceptual units.

However, the click-locating task, as defined in the early experiments, involved a complex interaction of perception and memory, since the subject had to remember the sentence he had just heard, remember where the click had occurred, and locate the click in a written version of the sentence.

Reaction time is a response measure that is more directly linked to perception in that the subject is not required to remember the click location. But when reaction time to clicks was measured, it was found that reaction time was not shortest to clicks located in constituent boundaries, as the theory would predict, and furthermore, reaction time did not seem to be related to the syntactic structure of a sentence (Abrams and Bever, 1970).

In order to explain this development, Abrams and Bever suggest a different model of attention in speech perception; they argue that the latency of the response to the click is a function of a subject's over-all attention to sensory input. At the beginning of a clause, the subject must pay attention to the input very closely, hence his reaction time to clicks is fast. At the end of clauses, the subject can already predict much of what is to come, so he does not have to pay much attention, and his reaction time to clicks is slower.

But it is also possible that constituent structure is not directly involved in perception, but is a result of perceptual analysis. It is possible that reaction time is a function of the suprasegmental structure of a sentence, as suggested by Dr. Lehiste (personal communication).

An experiment was designed to test a part of this hypothesis, namely to determine whether reaction time to clicks is affected by their relation to stressed elements.

Method

Stimuli: Ten sentences were selected to serve as stimuli. Each sentence was recorded two times in random order. Sentences were separated by a pause of 5 seconds. The recording was made in a sound-proof booth and an Ampex 350 tape recorder, at 7 1/2 i.p.s.

The speaker was male, with a medium pitched voice. He was instructed to say the sentences clearly and naturally.

One click was placed in each sentence. There were four types of click location: in a stressed vowel, in an unstressed vowel, in the consonant preceding a stressed vowel, and in the consonant preceding an unstressed vowel. In addition, one click was located in a constituent boundary. The clicks were produced by a capacitor discharge, triggered by the release of a key. The click so produced was a single spike, with a very rapid rise and decay. The duration of each click was approximately 25 msec.

The stimulus tape was made by re-recording the sentences on one channel of an Ampex 354 tape recorder and recording the click, at the appropriate time, on the second channel. In addition, five clicks were recorded on the stimulus tape before the clicks which were associated with sentences, to determine each subject's reaction time to non-speech stimuli.

The sentences employed, and the location of the clicks, are given below. For convenience, the location of clicks in both productions of the sentence is shown in one written version of the sentences. The complex sentences are taken from the study conducted by Abrams and Bever (1970); the simple sentences are taken from a study conducted by Lehiste (1971).

1. That the matter was dealt with |fast, |was a surprise
   to Harry.
2. Since she was free |that day, |her friends asked her to come.

3. My sleep was disturbed.
4. By making his plan known, Jim brought out the objections of everybody.
5. Speed kills.
6. Any student who is bright but young, would not have seen it.
7. The speed was controlled.
8. Sleep refreshes.
9. If you did call up Bill, I thank you for your trouble.
10. After the dry summer of that year, some of the crops were completely lost.

Click location was verified by inspecting the oscillograms, produced by two channels of an Elema-Schönander Mingograf, representing the two channels of the stimulus tape.

Subjects: Eleven subjects participated in the experiment. All were members of the Ohio State University linguistics department.

Procedure: Each subject listened to the stimulus tape two times. The first time, he was instructed to listen to the sentences and to push a key as quickly as he could when he heard the click. The key triggered a capacitor discharge which was recorded directly on one channel of an Elema-Schönander Mingograf. Simultaneously, the channel of the stimulus tape which contained the clicks was recorded on another channel of the Mingograf. The instrumentation is shown in the accompanying diagram (Fig. 9). Paper speed was 100 mm per second.

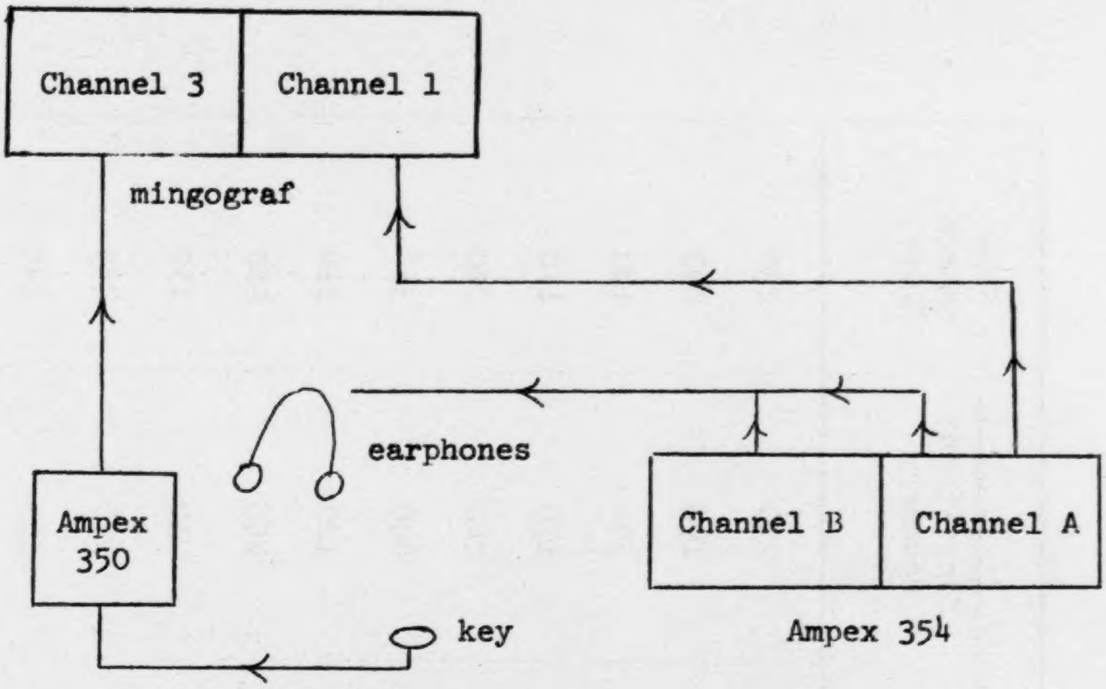


Fig. 9. Instrumentation for "click" experiment.

Immediately after the first test, the subject listened to the tape again. This time, he was provided with a written copy of each sentence and asked to mark the location of each click.

Reaction time to clicks was determined by measuring from the peak of the stimulus click to the onset of the response.

Results

The reaction time to clicks was compared for four conditions: when the click occurred in a stressed vowel, when it occurred in an unstressed vowel, when it occurred in a consonant preceding a stressed vowel, and when it occurred in a consonant preceding an unstressed vowel. The results are presented in Table 54 and in Fig. 10 to 12. Fig. 10 shows the reaction time to a click embedded in a consonant preceding a stressed vowel,and in a consonant preceding an unstressed vowel. For all but one subject, the reaction time is faster to the click preceding an unstressed vowel. Fig. 11 shows reaction time to

TABLE 54
MEAN REACTION TIME TO CLICKS (IN MSEC.)

	Click Location					Non-speech Click
Subject	Stressed Vowel	Consonant Preceding Stressed Vowel	Unstressed Vowel	Consonant Preceding Unstressed Vowel	Constituent Boundary	
1	333	336	293	278	230	230
2	276	242	226	195	190	200
3	235	244	237	218	390	180
4	255	241	198	150	170	170
5	233	239	249	240	240	200
6	524	557	582	505	600	390
7	236	181	144	145	120	110
8	406	383	316	313	460	280
9	241	283	206	200	210	150
10	236	230	224	220	250	235
11	163	165	175	175	165	140
For all subjects	285	281	259	240	275	

clicks embedded in stressed vowels and to clicks embedded in unstressed vowels. For six subjects, the reaction time is faster to a click in an unstressed vowel; for the other subjects, the reaction time is essentially the same.

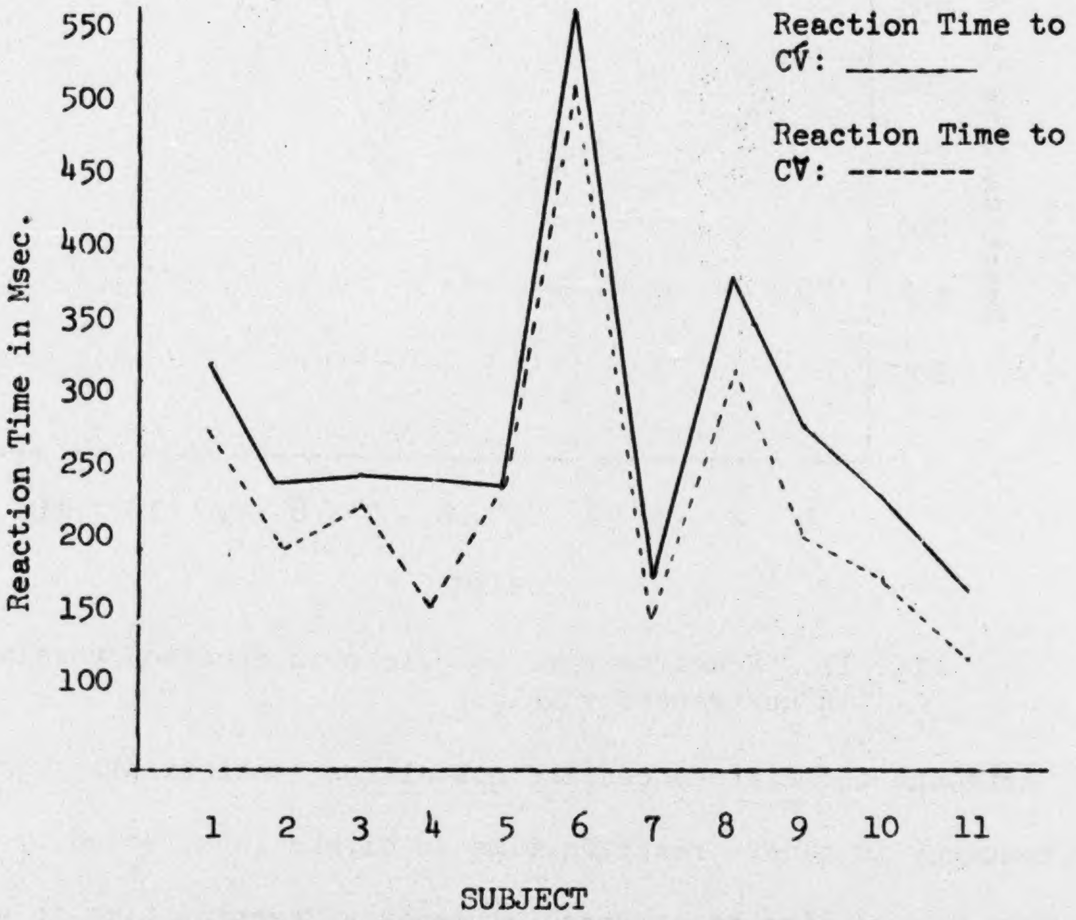


Fig. 10. Reaction time to clicks in consonants preceding stressed vowels and to clicks in consonants preceding unstressed vowels.

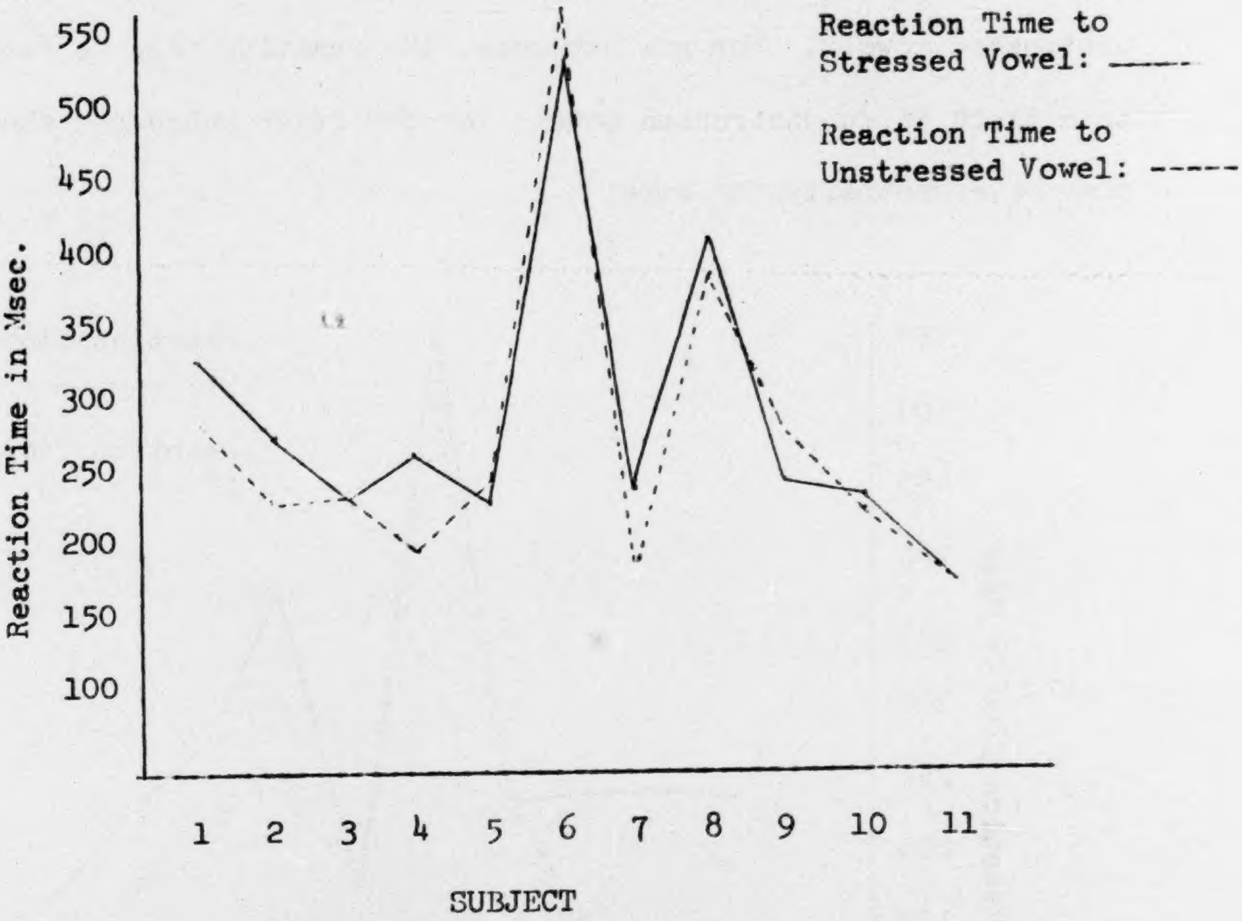


Fig. 11. Reaction time to clicks in stressed vowels and in unstressed vowels.

Although the differences are not always statistically significant, the tendency is clear: reaction time to clicks is affected by their location in relation to stressed elements. Reaction time to a click is longest when the click is in the vicinity of a stressed element, either in a stressed vowel or in a consonant preceding a stressed vowel. Reaction time is shorter when the click is in the vicinity of an unstressed element, either in an unstressed vowel or in a consonant preceding an unstressed vowel.

The reaction time to clicks located in constituent boundaries is quite variable. For some subjects, it is very short in this condition, approaching the reaction time to non-speech stimuli. For other subjects, it is quite long, longer than the reaction time to clicks in any other condition.

Reaction time to non-speech clicks is short in all cases, implying that reacting to a click in a speech context is more complex than simply reacting to a click. These results are presented in Fig. 12.

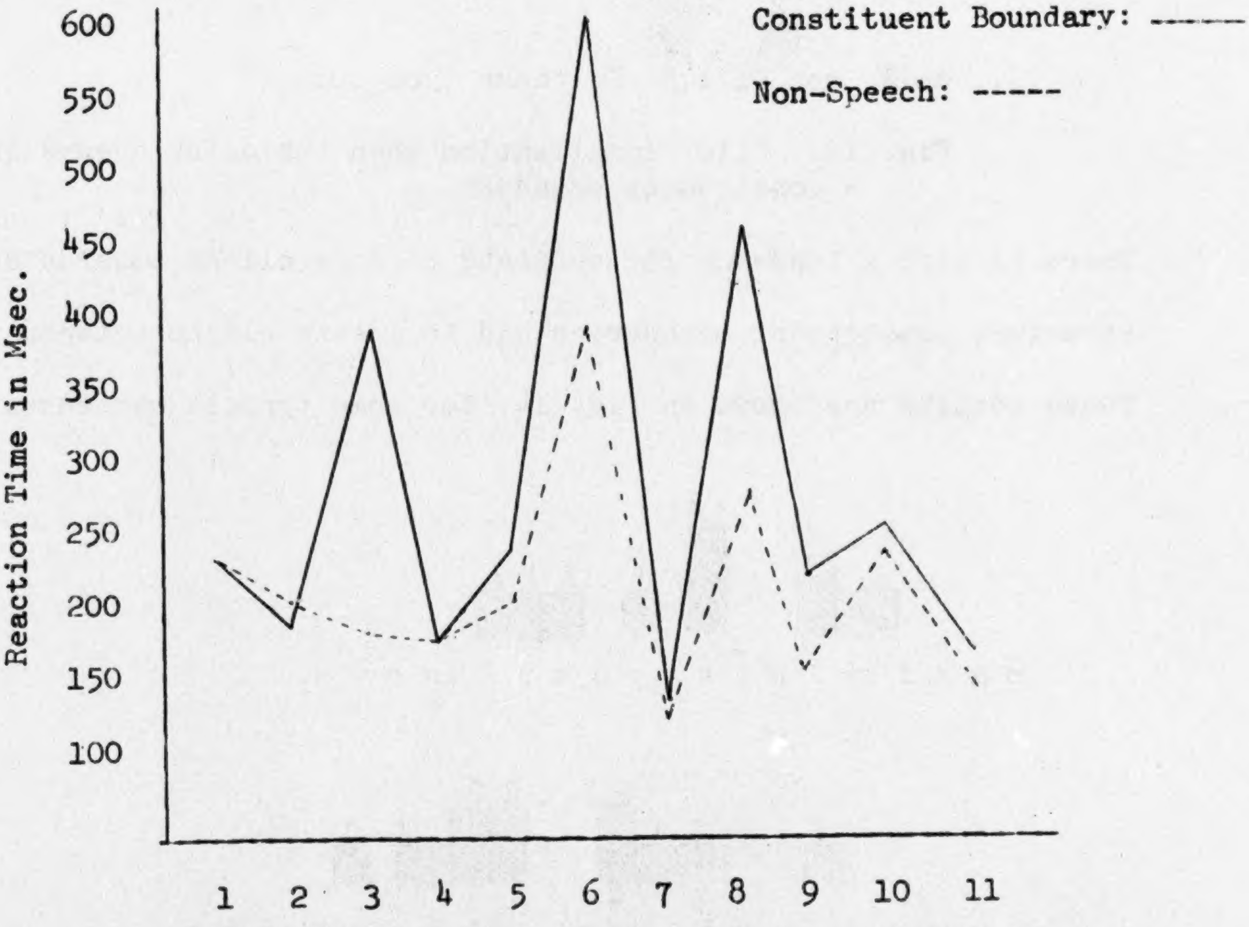


Fig. 12. Simple reaction time to click, and reaction time to click in a constituent boundary.

There is considerable variation in reaction time between subjects: subject 6, particularly, has quite slow reaction time to all conditions. Nevertheless, for each subject, the reaction times are in the same relationships, depending on the location of the click.

Click localization: The results of the click localization test are, in general, in agreement with previous studies. Click localization tends to be accurate when the click occurs in a constituent boundary. This is shown in Fig. 13. The asterisk indicates the location of the click; the bar graph indicates the subjects' localization of the click.

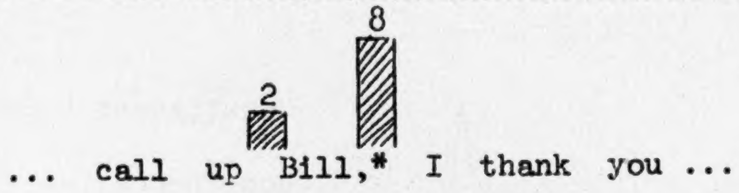
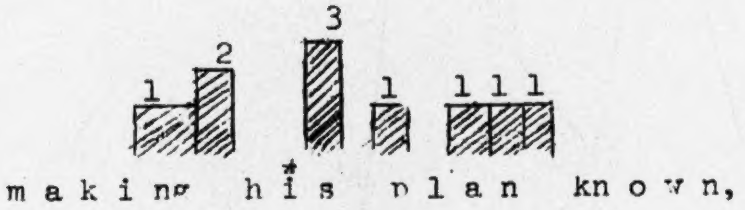
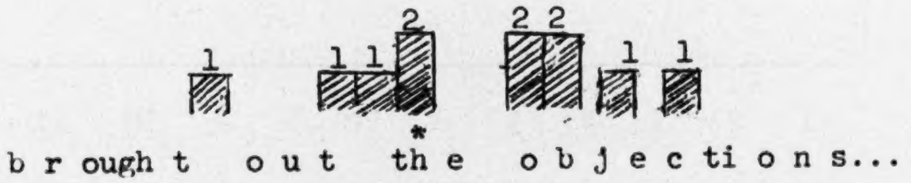


Fig. 13. Click localization when the click occurs in a constituent boundary.

There is also a tendency for subjects to move clicks towards deep structure constituent boundaries and to locate clicks between words. These results are shown in Fig. 14, for some typical sentences.





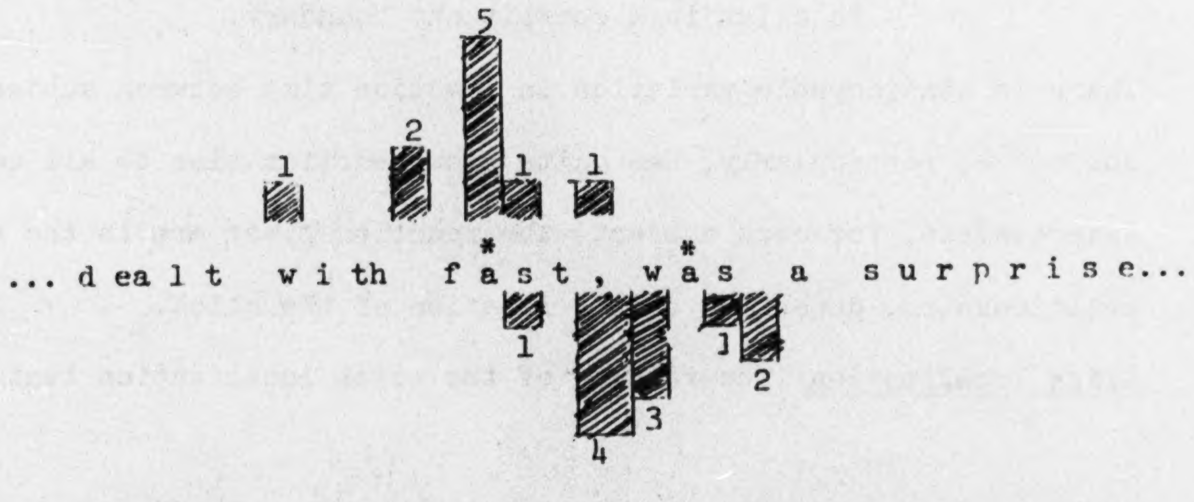


Fig. 14--continued

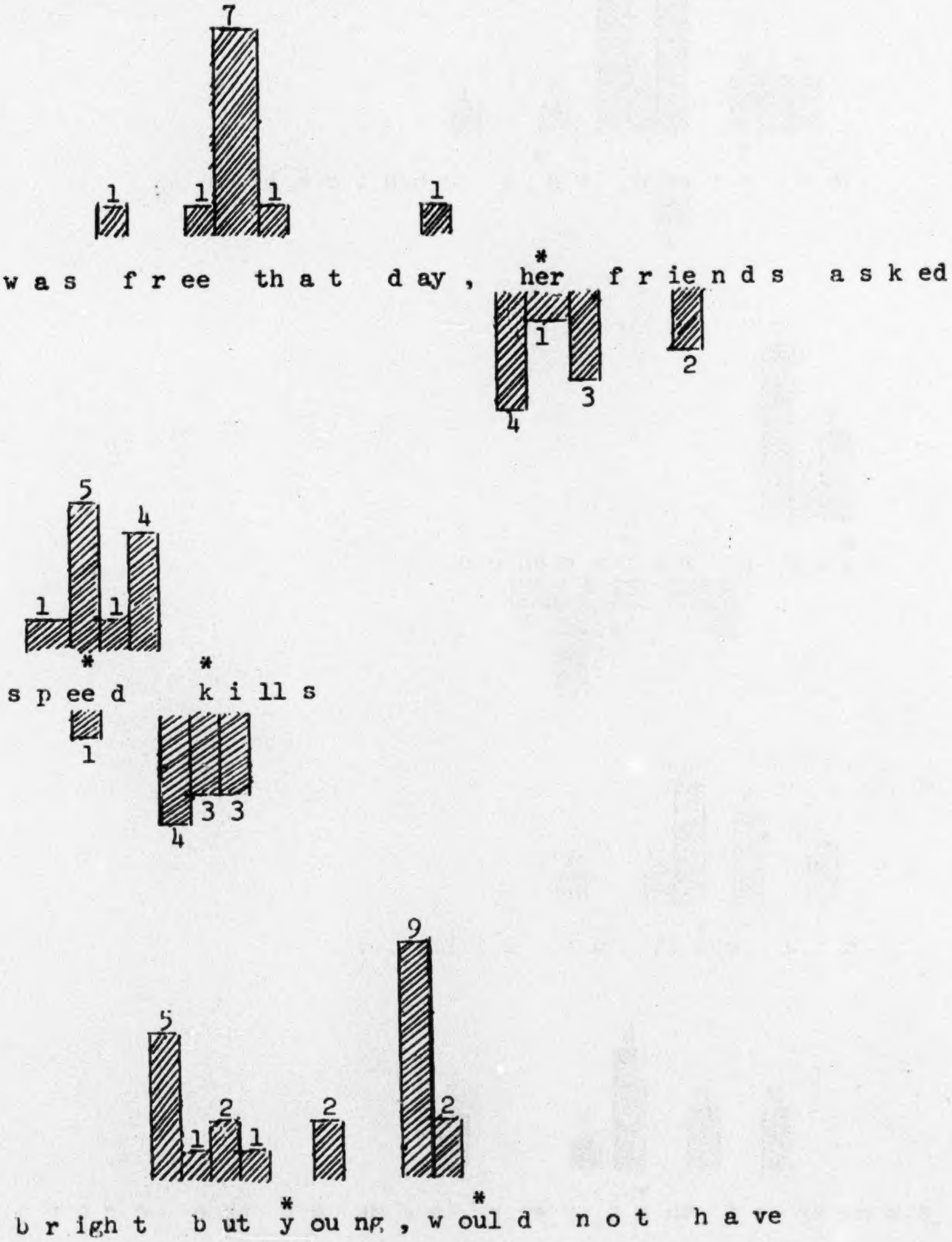


Fig. 14--continued

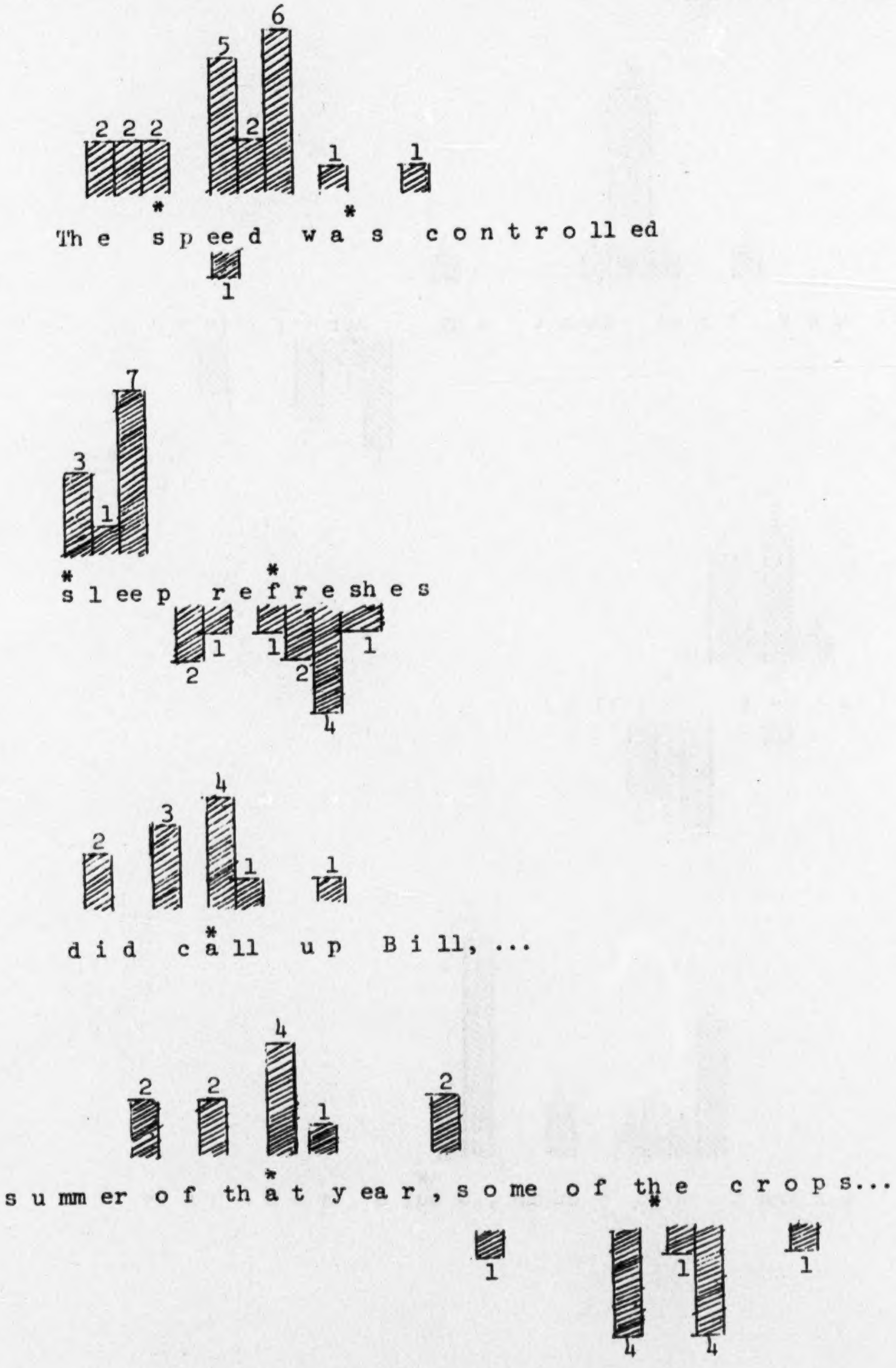


Fig. 14. Click localization.

However, the location of stress also affects click localization. Clicks in stressed vowels are localized much more accurately than clicks in unstressed vowels. This can be clearly seen by examining Fig. 15. The click in the stressed vowel of sleep is localized correctly more often than the click in the unstressed vowel of was. Furthermore, subjects do not miss the correct location by as much for the click in the stressed vowel as for the click in the unstressed vowel.

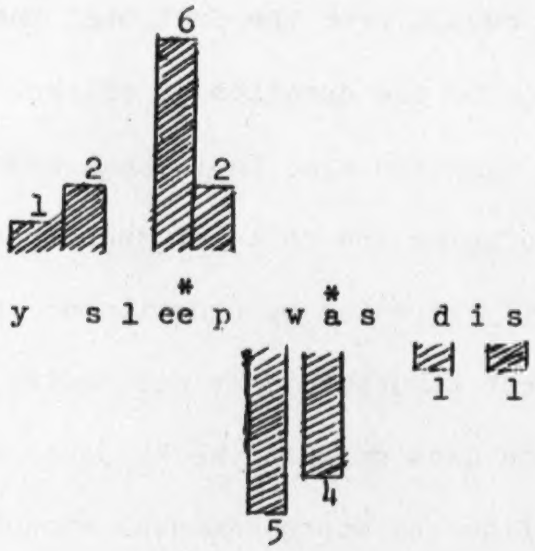


Fig. 15. Click localization in stressed and unstressed vowels.

Accuracy of click localization is summarized in Table 55.

TABLE 55

CLICK LOCALIZATION: PER CENT CORRECT

Stressed vowel	Unstressed vowel	Consonant	Constituent boundary
46	12	12	81

Discussion

The click localization data seem to imply that click localization is controlled by two parameters, constituent structure and the presence of stress. Click localization errors tend to lie in the direction predicted by theory, but clicks are less likely to be moved from a stressed vowel than from an unstressed vowel. That localization of clicks in consonants is also inaccurate may simply be a result of response bias: subjects may be less inclined to locate a click in a consonant. However, it may also result from the fact that the duration of consonants is short in relation to the duration of clicks.

The observed differences in reaction time imply that suprasegmental structure has some function in defining the units of speech perception. Since reaction time is not directly affected by constituent structure, it can be inferred that constituent structure does not define the units of perceptual input. Instead, the data support the hypothesis that units of perceptual input are defined by suprasegmental structure, i.e. stress and intonation.

There is one objection that might be raised to this conclusion. Stressed vowels occur in words that have semantic content whereas unstressed vowels occur in words that have less semantic content. In other words, words with stressed vowels are not predictable from context while words with unstressed vowels are much more readily predictable. The experiment, as designed, does not explicitly differentiate between this effect and the presence of stress. However, the objection is not crucial because the effect on reaction time is quite as pronounced when the click is in the consonant preceding the vowel. It is difficult to see why a subject should react differently to these clicks if only the predictability of the word were the issue. Further testing is necessary, however, to rule out the "predictability hypothesis" completely.

CHAPTER FIVE

CONCLUSION

The results of the studies reported above are interesting in themselves, but they are also interesting in what they imply about the processes underlying speech perception. To summarize briefly, the results are the following:

1. Subjects are aware of sub-phonemic phonetic differences, at least under appropriate conditions, but can not make linguistic use of them.

2. Perception of at least some phonological segments involves special perceptual mechanisms, rather than proceeding segment-by-segment.

3. Syntactic units in perception may be defined by suprasegmental structure.

The Need for Perceptual Units

Before the implications of these findings for specific theories of speech perception will be discussed, it seems reasonable to re-examine the assumption of this study, namely that there are units in speech perception.

As Experiment I shows, subjects can become aware of very fine phonetic differences if they attend to a particular utterance with great care. It is likely that subjects could even be taught to identify most of the words used in Experiment I properly, provided that the subjects got proper feedback, and provided that the stimuli were properly selected so that the distinctive cues were invariably present in each production. In this sense, there is no clear lower limit below which speech stimuli are perceived as "the same," and, one might suppose, no lower limit for a phonological perceptual unit either.

However, just because a listener can utilize fine phonetic detail when the conditions of a test force him to do so, does not imply that listeners inevitably notice or pay attention to such information. Rather, listeners are probably content with less detailed phonetic representations. To draw an analogy with visual perception, we do not examine leaves when we are looking at a forest. In visual perception, we can examine, in great detail, the shape and color of particular objects. But ordinarily, we do not do this; we are content to recognize objects and to behave appropriately to them--we sit in chairs, pat dogs, speak to our friends. Similarly, in the ordinary course of language use, we deal with something other than with fine phonetic differences. Therefore, there must be postulated some larger unit--or higher level--at which the phonological structure of an utterance is represented, independently of the fine phonetic details of the utterance.

This level, however, must be independent of syntactic or contextual information for the reason that new words, such as proper names and technical terms, do not present undue difficulty to us; we simply hear the word, and we remember it.

These two considerations imply a lower and an upper boundary for the perception and coding of phonological information: the units involved in this process can not be equivalent to the phonetic representation of the utterance and the units can not be dependent on syntactic information.

Similarly, there must be some unit, or preferred units, in arriving at a syntactic analysis of a sentence. It is not possible for listeners to store a whole sentence in memory, simply because, unless the sentence were recoded in some way, it would very easily exceed the short-term memory capacity of a listener. It seems reasonable to suppose that the recoding operation can not process the sentence continuously as it is heard, but that the sentence must be broken up into some sort of units--perceptual segmentation units--for the recoding process to operate upon. The results of the recoding process certainly embody syntactic structure in some way.

It has been supposed previously that the perceptual segmentation units were syntactic as well. But the results of Experiment III can not be reconciled with the idea that segmentation units are syntactic. If they were, then reaction time to clicks and click localization should give the same results. Since this is not the case, the implication is that, at some level, sentences are processed in terms of non-syntactic units. The results of Experiment III imply that these units are defined by the phonological structure of an utterance and that these units function at the initial segmentation of the sentence. These initially segmented units are then recoded, probably by assigning them a particular syntactic function.

Thus, there is a need for at least two types of units in speech perception: units of phonological processing and units defining a part of a sentence for further syntactic analysis--perceptual segmentation units.

Implications for Perception Models

Not all of the theories of speech perception discussed in Chapter I make specific predictions about units of speech perception, but several do, namely the motor theory, analysis-by-synthesis, "filtering" theories, Osgood's perception model, and the perceptual strategies model. The experimental findings, reported above, conflict with some predictions made by these models, although, of course, the models may be revised slightly to cope with them.

First, the motor theory of speech perception, in that it asserts categorial perception of phonemes, conflicts with a listener's ability to become aware of sub-phonemic phonetic differences. If the perception of phonemes were indeed categorial, then listeners could not become aware of any sub-phonemic information whatever. Yet this is not the case; listeners are aware of sub-phonemic detail and use both vowel length and consonant quality in developing a strategy for making identification judgments. Second, that the motor theory postulates a phoneme-like unit as the basic unit of perception, it conflicts with the implications of Experiment II--that listeners apparently employ special perceptual mechanisms to process some consonant clusters, rather than perceiving the clusters "phoneme-by-phoneme."

This second objection also applies to analysis-by-synthesis models. These models assume that phonology is perceived in terms of discrete segments. This assumption can not account for the finding of Experiment II--that reversal of the order of segments is the most common perceptual error.

In a fundamental way, the motor theory and analysis-by-synthesis are quite similar: both postulate that the listener generates a possible phonetic output and matches this output against the incoming message. The theories differ only in the nature of the internal mechanisms that they postulate. The experiments reported in this work do not have any implications for this basic postulate. However, it must be added here that there is no evidence that such internal mechanisms are strictly necessary. The "synthesis" theories have been postulated, apparently, because there are no invariants given immediately in the acoustic speech signal. Instead, the relationship between the acoustic signal and the perceptual result is quite complex.

Still, this difficulty is not unique to speech perception. In the study of visual perception, it has been commonly observed that the retinal image--which we may consider to be analogous to the acoustic input--is much more varied than the perception of objects. The retinal image changes radically as we view an object from different angles and from different distances, yet the percept is of an unchanging, stable object. The relationship between the retinal image and the percept is no less complex than the relationship between the acoustic signal and perceived speech, yet we do not posit a "motor theory of visual perception" for this reason.

These comments are added only to point out that a complex relationship is not sufficient grounds for positing intermediate devices of an unrelated type: theoretical mechanisms have to have independent empirical justification.

The filtering theories discussed in Chapter I are of two types: theories that assume a phoneme-like unit, and Wickelgren's context-sensitive coding which assumes that the perceptual unit is similar to the traditional allophone. There are two objections to the phoneme-like unit: first, the well-known lack of invariance between phonemes and the acoustic signal and, second, the fact that obstruent clusters are apparently not perceived "phoneme-by-phoneme."

Wickelgren's theory tries to overcome the first difficulty by assuming smaller, hence presumably invariant, units, but it does so at the cost of proliferating the number of different units that must be assumed. Furthermore, it is still to be determined if there are invariant acoustic differences that can be used to determine the order of segments. Context-sensitive coding can, however, account for the perception of obstruent clusters. One further advantage of both types of filtering theories must be mentioned. Neither version of the theories is limited to a strict sequence of segments in the input, if the "filters" can be assumed to be working in parallel. Rather, the listener can be presumed to process a rather large segment of speech at one time.

Osgood argues that the word is the basic perceptual unit. However, there are several difficulties with this position. First, as has already been pointed out, there must be some perceptual units which enable a listener to code a new word. It would be unparsimonious to suppose that these mechanisms are used only to code new words. Second, listeners can become aware of very subtle phonetic differences, a finding which is counter to the notion that a word is the only perceptual unit. But it seems likely that words function as units at some level of speech perception.

The perception of syntactic structure has been touched on only briefly in this study. The perceptual strategies suggested by Bever, and others, are not in dispute here; a fair amount of evidence has been offered to substantiate them, and no finding presented in this work conflicts with them. What has been questioned is the assumption that syntactic units provide the initial segmentation of a sentence. As has already been pointed out, this can not be the case because reaction time and click localization do not give the same results. Rather, the most likely hypothesis is that initial segmentation is accomplished by using the suprasegmental structure of an utterance. After this initial segmentation, perceptual strategies, as defined by Bever, may well apply to enable the listener to arrive at a syntactic analysis of the utterance.

The remarkable fact about speech perception is that it seems to be an easy and effortless process. Yet the mechanisms underlying this process are only beginning to be studied. Perhaps the best that could be said is that we are beginning to appreciate how complicated and mysterious the process of speech perception really is. Any adequate explanation will undoubtedly require a much more thorough understanding of human cognitive abilities on the one hand, and of the nature of language on the other.

BIBLIOGRAPHY

Abrams, K., and T. G. Bever, "Syntactic Structure Modifies Attention during Speech Perception and Recognition," Quarterly Journal of Experimental Psychology 21 (1969), 280-290.

Abramson, A. S., "Identification and Discrimination of Phonemic Tones," Journal of the Acoustical Society of America 33 (1961),8 842.

Abramson, A. S., and L. Lisker, "Discriminability along the Voicing Continuum: Cross-language Tests," Proceedings of the 6th International Congress of Phonetic Sciences, Prague (1967).

Bastian, J., and A. S. Abramson, "Identification and Discrimination of Phonemic Vowel Duration," Journal of the Acoustical Society of America 34 (1962), 743-644.

Bastian, J., P. Delattre, and A. M. Liberman, "Silent Interval as a Cue for the Distinction between Stops and Semivowels in Medial Position," Journal of the Acoustical Society of America 31 (1959), 1568.

Bastian, J., P. D. Eimas, and A. M. Liberman, "Identification and Discrimination of a Phonemic Contrast Induced by Silent Interval," Journal of the Acoustical Society of America 33 (1961), 842.

Bever, T. G., "The Cognitive Basis for Linguistic Structures," in Hayes, J., (ed.), Cognition and the Development of Language, New York: Wiley (1970).

Bever, T. G., J. Lackner, and R. Kirk, "The Underlying Structure Sentence is the Primary Unit of Speech Perception," Perception and Psychophysics 5 (1969), 225-234.

Bever, T. G. and D. T. Langendoen, "The Interaction of Speech Perception and Grammatical Structure in the Evolution of Language," unpublished manuscript.

Blesser, B. A., "Inadequacy of a Spectral Description in Relationship to Speech Perception," Paper presented at the 78th meeting of the Acoustical Society of America (4-7 November 1969).

Bloomfield, Leonard, *Language*, New York (1933).

Blumenthal, A., "Prompted Recall of Sentences," *Journal of Verbal Learning and Verbal Behavior* 6 (1967), 203-206.

Bondarko, L. V., N. G. Zagorujko, V. A. Koževnikov, A. P. Molčanov, and L. A. Čistovič, "A Model of Speech Perception by Humans," (Ilse Lehiste, translator), *Working Papers in Linguistics* 6 (The Ohio State University, 1970), 89-132.

Broadbent, D. E., and M. Gregory, "Accuracy of Recognition for Speech Presented to the Right and Left Ears," *Quarterly Journal of Experimental Psychology* 16 (1964), 359-60.

Broadbent, D. E., and P. Ladefoged, "Auditory Perception of Temporal Order," *Journal of the Acoustical Society of America* 31 (1959) , 1539.

Bryden, M. P., "Ear Preferences in Auditory Perception," *Journal of Experimental Psychology* 65 (1963), 103-105.

Clark, H. H. and E. V. Clark, "Semantic Distinctions and Memory for Complex Sentences," *Quarterly Journal of Experimental Psychology* 20 (1968), 129-138.

Clark, H. H. and R. A. Stafford, "Memory for Semantic Features," *Journal of Experimental Psychology* 80 (1969), 326-334.

Cooper, F. S., "Research on Reading Machines for the Blind," in P. A. Zahl (ed.), Blindness: Modern Approaches to the Unseen Environment, Princeton Univ. Press (1950).

Cooper, F. S., "Describing the Speech Process in Motor Command Terms," Journal of the Acoustical Society of America 39 (1966), 1221 A.

Cooper, F. S., P. C. Delattre, A. M. Liberman, J. M. Borst, and H. G. Gerstman, "Some Experiments on the Perception of Synthetic Speech Sounds," Journal of the Acoustical Society of America 24 (1952), 597-606.

Cooper, F. S., A. M. Liberman, and J. M. Borst, "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech," Proceedings of the National Academy of Sciences 37 (1951), 318-325.

Cross, D. V. and H. L. Lane, "On the Discriminative Control of Concurrent Responses: The Relations among Response Frequency, Latency, and Topography in Auditory Generalization," Journal of the Experimental Analysis of Behavior 5 (1962), 487-496.

Cross, D. V., H. L. Lane, and W. C. Sheppard, "Identification and Discrimination Functions for a Visual Continuum and Their Relation to the Motor Theory of Speech Perception," Journal of Experimental Psychology 70 (1965), 63-74.

Davis, H., "Auditory Communication," Journal of Speech and Hearing Disorders 16 (1951), 3-8.

Delattre, P. C., A. M. Liberman, and F. S. Cooper, "Acoustic Loci and Transitional Cues for Consonants," Journal of the Acoustical Society of America 27 (1955), 679-773.

Delattre, P., A. M. Liberman, F. S. Cooper, and L. G. Gerstman, "An Experimental Study of the Acoustic Determinants of Vowel Color," Word 8 (1952), 195-210.

Denes, P.,"Effect of Duration on the Perception of Voicing," Journal of the Acoustical Society of America 27 (1955), 761-766.

Denes, P., "On the Motor Theory of Speech Perception," Proceedings of the 5th International Congress of Phonetic Sciences, (Münster, Basel, New York: S. Karger, 1964) 232-238.

Dixon, Theodore R., and David L. Horton (eds.), Verbal Behavior and General Behavior Theory (Englewood Cliffs., N.J.: Prentice-Hall, 1968).

Eimas, P., "The Relation between Identification and Discrimination along Speech and Non-speech Continua," Language and Speech 6 (1963), 206-217.

Fairbanks, G., "A Theory of the Speech-Mechanism as a Servo-system," Journal of Speech and Hearing Disorders 19 (1954), 133-139.

Flanagan, J. H., "Difference Limen for Vowel Formant Frequency," Journal of the Acoustical Society of America 27 (1955), 613-617.

Fodor, J. A., and T. G. Bever, "The Psychological Reality of Linguistic Segments," Journal of Verbal Learning and Verbal Behavior 4 (1965), 414-420.

Fodor, J., and M. Garrett, "Some Syntactic Determinants of Sentential Complexity," Perception and Psychophysics 2 (1967), 289-296.

Fodor, J., M. Garrett, and T. Bever, "Some Syntactic Determinants of Sentential Complexity II: Verb Structure," Perception and Psychophysics 3 (1968), 453-461.

Fromkin, V. A., "Neuro-muscular Specification of Linguistic Units," *Language and Speech* 9 (1966), 170-199.

Fry, D. B., "Duration and Intensity as Physical Correlates of Linguistic Stress," *Journal of the Acoustical Society of America* 27 (1955), 765-768.

Fry, D. B., "Perception and Recognition in Speech," in *For Roman Jakobson*, M. Halle (ed.), (The Hague: Mouton & Co., 1956).

Fry, D. B., "Experimental Evidence for the Phoneme," in *In Honor of Daniel Jones*, D. Abercrombie, et al. (eds.), (London: Longmans, 1964).

Fry, D. B., "The Function of the Syllable," *Zeitschrift für Phonetik* 17 (1964), 215-221.

Fry, D. B., "Reaction Time Experiments in the Study of Speech Processing," *Progress Report*, Phonetics Laboratory, University College, London (1968).

Fry, D. B., A. S. Abramson, P. D. Eimas, and A. M. Liberman, "The Identification and Discrimination of Synthetic Vowels," *Language and Speech* 5 (1962), 171-189.

Fry, D. B. and P. Denes, "An Analogue of the Speech Recognition Process," *Mechanisation of Thought* (London, 1958).

Garrett, M., T. G. Bever, and J. A. Fodor, "The Active Use of Grammar in Speech Perception," *Perception and Psychophysics* 1 (1966), 30-32.

Gibson, Eleanor J., *Principles of Perceptual Learning and Development*, (Appleton-Century-Crofts: New York, 1969).

Gibson, James J., *The Senses Considered as Perceptual Systems*, (Houghton Mifflin: Boston, 1966).

Greenberg, J. H., and J. J. Jenkins, "Studies in the Psychological Correlates of the Sound System of American English," Word (1964), 157-177.

Halle, M., G. W. Hughes, and J.-P. A. Radley, "Acoustic Properties of Stop Consonants," Journal of the Acoustical Society of America 29 (1957), 107-116.

Halle, M., and K. N. Stevens, "Speech Recognition: A Model and a Program for Research," in The Structure of Language: Readings in the Philosophy of Language, J. A. Fodor and J. J. Katz, (eds.), (Englewood Cliffs, N.J.: Prentice-Hall, 1964).

Harris, K., "Cues for the Discrimination of American English Fricatives in Spoken Syllables," Language and Speech 1 (1958), 1-7.

Harris, K., J. Bastian, and A. M. Liberman, "Mimicry and the Perception of a Phonemic Contrast Induced by Silent Interval: Electromyographic and Acoustic Measures," Journal of the Acoustical Society of America 33 (1961), 842.A.

Harris, K., H. Hoffman, A. M. Liberman, P. C. Delattre, and F. S. Cooper, "Effect of Third-formant Transitions in the Perception of the Voiced Stop Consonants," Journal of the Acoustical Society of America 30 (1958), 122-126.

Hirsch, I. J., "Auditory Perception of Temporal Order," Journal of the Acoustical Society of America 31 (1959), 759-767.

Hockett, C., A Manual of Phonology, Indiana University Publications in Anthropology 17 (1955).

Hoffman, Howard S., "A Study of Some Cues in the Perception of the Voiced Stop Consonants," Journal of the Acoustical Society of America 30 (1958), 1035-1041.

House, A. S. and G. Fairbanks, "The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels," Journal of the Acoustical Society of America 25 (1953), 105-113.

House, A. S., K. N. Stevens, T. Sandel, and G. Arnold, "On the Learning of Speech-like Vocabularies," Journal of Verbal Learning and Verbal Behavior 1 (1962), 133-143.

Hughes, G. W. and M. Halle, "Spectral Properties of Fricative Consonants," Journal of the Acoustical Society of America 28 (1956), 303-310.

Jakobovits, L. A., and M. S. Miron (eds.), Readings in the Psychology of Language, (Englewood Cliffs, N.J.: Prentice-Hall, 1967).

Johnson, N., "The Psychological Reality of Phrase-structure Rules," Journal of Verbal Learning and Verbal Behavior 5 (1966), 469-475.

Joos, M., Acoustic Phonetics, Supplement to Language 24 (1948).

Katz, G., "Mentalism in Linguistics," Language 40 (1964), 124-138.

Kozhevnikov, V. A. and L. A. Chistovich, Speech: Articulation and Perception. U. S. Department of Commerce, JPRS Report 30,543 (1965).

Ladefoged, P., "The Perception of Speech," in Mechanisation of Thought Processes, VI (London, 1959).

Ladefoged, P., Three Areas of Experimental Phonetics, (London: Oxford University Press, 1967).

Ladefoged, P., and D. E. Broadbent, "Information Conveyed by Vowels," Journal of the Acoustical Society of America 29 (1957), 98.

Ladefoged, P., and D. E. Broadbent, "Perception of Sequence in Auditory Events," Quarterly Journal of Experimental Psychology 12 (1960) 162-70.

Lane, H. L., "Psychophysical Parameters of Vowel Perception," Psychological Monographs 76 (1962), 44.

Lane, H. L., "The Motor Theory of Speech Perception: A Critical Review," Psychological Review 72 (1965), 275-309.

Lehiste, I., Suprasegmentals (Cambridge, Mass.: M.I.T. Press, 1970).

Lehiste, I., "The Temporal Realization of Morphological and Syntactic Boundaries," Paper presented at the 81st Meeting of the Acoustical Society of America (April, 1971).

Lehiste, I. and G. E. Peterson, "Vowel Amplitude and Phonemic Stress in American English," Journal of the Acoustical Society of America 31 (1959), 428-435.

Liberman, A. M., "Some Results of Research on Speech Perception," Journal of the Acoustical Society of America 29 (1957), 117-123.

Liberman, A. M., F. S. Cooper, K. S. Harris, and P. F. MacNeilage, "A Motor Theory of Speech Perception," Proceedings of the Speech Communication Seminar (Stockholm: Royal Institute of Technology, 1962).

Liberman, A. M., F. S. Cooper, K. S. Harris, P. F. MacNeilage, and M. Studdert-Kennedy, "Some Observations on a Model for Speech Perception," Models for the Perception of Speech and Visual Form (M.I.T. Press, 1965).

Liberman, A. M., F. S. Cooper, D. S. Shankweiler, and M. Studdert-Kennedy, "Perception of the Speech Code," Psychological Review 74 (1967), 431-461.

Liberman, A. M., P. C. Delattre, and F. S. Cooper, "The Role of Selected Stimulus Variables in the Perception of the Unvoiced Stop Consonants," American Journal of Psychology 65 (1952), 497-516.

Liberman, A. M., P. C. Delattre, and F. S. Cooper, "Some Cues for the Distinction between Voiced and Voiceless Stops in Initial Positions," Language and Speech 1 (1958), 153-167.

Liberman, A. M., P. C. Delattre, F. S. Cooper, and H. J. Gerstman, "The Role of Consonant-vowel Transitions in the Perception of Stop and Nasal Consonants," Psychological Monograph 68 (1954).

Liberman, A. M., P. C. Delattre, H. J. Gerstman, and F. S. Cooper, "Tempo of Frequency Change as a Cue for Distinguishing Classes of Speech Sounds," Journal of Experimental Psychology 52 (1956), 127-137.

Liberman, A. M., K. S. Harris, P. Eimas, L. Lisker, and J. Bastian, "An Effect of Learning on Speech Perception: The Discrimination of Durations of Silence with and without Phonemic Significance," Language and Speech 4 (1961), 175-195.

Liberman, A. M., K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The Discrimination of Speech Sounds within and across Phoneme Boundaries," Journal of Experimental Psychology 54 (1957), 358-367.

Liberman, A. M., K. S. Harris, J. Kinney, and H. L. Lane, "The Discrimination of Relative Onset Time of the Components of Certain Speech and Non-speech Patterns," Journal of Experimental Psychology 61 (1961), 379-388.

Licklider, J. C. R., "On the Process of Speech Perception," Journal of the Acoustical Society of America 24 (1952), 590-594.

Licklider, J. C. R., and G. A. Miller, "The Perception of Speech," in Handbook of Experimental Psychology, S. S. Stevens, ed., (New York: John Wiley & Sons, 1951).

Lieberman, P., "Some Effects of Semantic and Grammatical Context on the Production and Perception of Speech," Language and Speech 6 (1963), 172-187.

Lieberman, P., Intonation, Perception, and Language (Cambridge, Mass.: M.I.T. Press, 1967).

Lindgren, N., "Machine Recognition of Human Language," IEEE Spectrum (March, 1965), 114-136; (April, 1965), 45-59.

Lisker, L., "Minimal Cues for Separating /w,r,l,j/ in Intervocalic Position," Word 13 (1957), 257-267.

Lisker, L., "Closure Duration and the Intervocalic Voiced/Voiceless Distinction in English," Language 33 (1957), 42-49.

Lisker, L., "Anatomy of Unstressed Syllables," Journal of the Acoustical Society of America 30 (1958), 682, A.

Lisker, L., and A. S. Abramson, "Cross-language Study of Voicing in Initial Stops: Acoustical Measurements," Word 20 (1964), 384-422.

Lisker, L., and A. S. Abramson, "Stop Categories and Voice Onset Time," in Proceedings of the Fifth International Congress of Phonetic Sciences, (Münster, 1964. Basel: S. Karger, 1965).

Lisker, L., and A. S. Abramson, "The Voicing Dimension: Some Experiments in Comparative Phonetics," in Proceedings of the Sixth International Congress of Phonetic Sciences, (Prague, 1967).

Lotz, J., A. S. Abramson, H. Gerstman, F. Ingemann, and W. J. Nemser, "The Perception of English Stops by Speakers of English, Spanish, Hungarian, and Thai," Language and Speech 3 (1960), 71-77.

Lisker, L., F. S. Cooper, and A. M. Liberman, "The Uses of Experiment in Language Description," Word 18 (1962), 82-106.

Lyons, J., and R. G. Wales (eds.), Psycholinguistics Papers (Edinburgh: Edinburgh University Press, 1966).

Malécot, A., "Acoustic Cues for Nasal Consonants," Language 32 (1956), 274-284.

Mattingly, I., and A. M. Liberman, "The Speech Code and the Physiology of Language," in K. N. Leibovic, Information Processing in the Nervous System (New York: Springer-Verlag, 1969).

Mehler, J., "Some Effects of Grammatical Transformations on the Recall of English Sentences," Journal of Verbal Learning and Verbal Behavior 2 (1963), 346-351.

Mehler, J., and P. Carey, "Role of Surface and Base Structure in the Perception of Sentences," Journal of Verbal Learning and Verbal Behavior 6 (1967), 335-338.

Miller, G. A., "The Perception of Short Bursts of Noise," Journal of the Acoustical Society of America 20 (1948), 160-170.

Miller, G. A., "Speech and Language," in Handbook of Experimental Psychology, S. S. Stevens, ed., (New York: John Wiley & Sons, 1951).

Miller, G. A., "The Magical Number 7, Plus or Minus 2: Some Limits in Our Capacity for Processing Information," Psychological Review 63 (1950), 81-97.

Miller, G. A., "The Perception of Speech," in For Roman Jakobson: Essays on the Occasion of His 60th Birthday, M. Halle, ed., (The Hague: Mouton & Co., 1956).

Miller, G. A., "Speech and Communication," Journal of the Acoustical Society of America 30 (1958), 397-398.

Miller, G. A., "Some Psychological Studies of Grammar," American Psychologist 17 (1962), 748-762.

Miller, G. A., "Decision Units in the Perception of Speech," IRE Transactions on Information Theory IT-8 (1962), 81-83.

Miller, G. A., E. Galanter, and K. H. Pribram, Plans and the Structure of Behavior (New York: Henry Holt & Co., 1960).

Miller, G. A., G. A. Heise, and W. Lichten, "The Intelligibility of Speech as a Function of the Context of the Test Materials," Journal of Experimental Psychology 41 (1951), 329-335.

Miller, G. A., and S. Isard, "Some Perceptual Consequences of Linguistic Rules," Journal of Verbal Learning and Verbal Behavior 2 (1963), 217-228.

Miller, G. A., and P. E. Nicely, "An Analysis of Perceptual Confusions among Some English Consonants," Journal of the Acoustical Society of America 27 (1955), 338-352.

Mowrer, O. H., "The Psychologist Looks at Language," American Psychologist 9 (1954), 660-694.

Norman, D. A., Memory and Attention: An Introduction to Human Information Processing (New York: Wiley, 1969).

O'Connor, J. D., H. J. Gerstman, A. M. Liberman, P. S. Delattre, and F. S. Cooper, "Acoustic Cues for the Perception of Initial /w,j,r,l/ in English," Word 13 (1957), 24-43.

Öhman, S. E. G., "Coarticulation in VCV Utterances: Spectrographic Measurements," Journal of the Acoustical Society of America 39 (1966), 151-168.

Osgood, Charles E., "Psycholinguistics," in Psychology: A Study of a Science, S. Koch, ed. (New York: McGraw-Hill, 1963).

Osgood, Charles E., "On Understanding and Creating Sentences," American Psychologist 18 (1963), 735-751.

Peterson, G. E., "The Phonetic Value of Vowels," Language 27 (1951), 541-553.

Peterson, G. E., "The Information-bearing Elements of Speech," Journal of the Acoustical Society of America 24 (1952), 624-637.

Peterson, G. E., "Basic Physical Systems for Communication between Two Individuals," Journal of Speech and Hearing Disorders 18 (1953), 116-120.

Peterson, G. E., "An Oral Communication Model," Language 31 (1955), 414-427.

Peterson, G. E., and H. L. Barney, "Control Methods Used in a Study of the Vowels," Journal of the Acoustical Society of America 24 (1952), 175-184.

Schatz, C. D., "The Role of Context in the Perception of Stops," Language 30 (1954), 47-56.

Shankweiler, D., and M. Studdert-Kennedy, "An Analysis of Perceptual Confusions in Identification of Dichotically Presented CVC Syllables," Journal of the Acoustical Society of America 41 (1967), 1581.

Shankweiler, D., and M. Studdert-Kennedy, "Identification of Consonants and Vowels Presented to Left and Right Ears," Quarterly Journal of Experimental Psychology 19 (1967), 59-63.

Skinner, B. F., Verbal Behavior (New York: Appleton-Century-Crofts, 1957).

Stevens, K. N., "The Perception of Vowel Formants," Journal of the Acoustical Society of America 24 (1952), 450.

Stevens, K. N., "Toward a Model for Speech Recognition," Journal of the Acoustical Society of America 32 (1960), 47-55.

Stevens, K. N., and M. Halle, "Remarks on Analysis by Synthesis and Distinctive Features," Models for the Perception of Speech and Visual Form (Cambridge, Mass.: M.I.T. Press, 1965).

Stevens, K. N., A. M. Liberman, M. Studdert-Kennedy, and S. E. G. Öhman, "Cross-language Study of Vowel Perception," Language and Speech 12 (1969), 1-23.

Stevens, K. N., S. E. G. Öhman, and A. M. Liberman, "Identification and Discrimination of Rounded and Unrounded Vowels," Journal of the Acoustical Society of America 35 (1963), 1900.A.

Stolz, W. S., "A Study of the Ability to Decode Grammatically Novel Sentences," Journal of Verbal Learning and Verbal Behavior 6 (1967), 867-873.

Strevens, P., "Spectra of Fricative Noise in Human Speech," Language and Speech 3 (1960), 32-48.

Studdert-Kennedy, M., A. M. Liberman, K. S. Harris, and F. S. Cooper, "Motor Theory of Speech Perception: A Reply to Lane's Critical Review," Psychological Review 77 (1970), 234-249.

Studdert-Kennedy, M., A. M. Liberman, and K. N. Stevens, "Reaction Time to Synthetic Stop Consonants and Vowels at Phoneme Centers and at Phoneme Boundaries," Journal of the Acoustical Society of America 35 (1963), 1900. A.

Uldall, E., "Transitions in Fricative Noise," Language and Speech 7 (1964), 13-15.

Warren, R. M., and R. P. Warren, "Auditory Illusions and Confusions," Scientific American (December, 1970), 30-36.

Watson, John B., Behaviorism (New York: W. W. Norton, 1930).

Wickelgren, Wayne A., "Distinctive Features and Errors in Short-Term Memory for English Consonants," Journal of the Acoustical Society of America 39 (1966), 388-398.

Wickelgren, W. A., "Context-sensitive Coding, Associative Memory, and Serial Order in (Speech) Behavior," Psychological Review 76 (1969a), 1-15.

Wickelgren, W. A., "Context-sensitive Coding in Speech Recognition, Articulation, and Development," in Information Processing in the Nervous System, K. N. Leibovic, ed. (New York: Springer-Verlag, 1969b).

Wright, C. W. An English Word Count (Pretoria: National Bureau of Educational and Social Research, 1965).

The Temporal Realization of Morphological and Syntactic Boundaries*

Ilse Lehiste

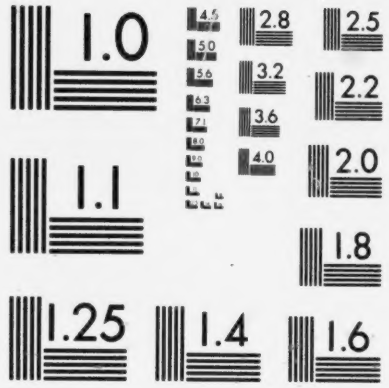
---

*Sponsored in part by the National Science Foundation through Grant GN-534.1 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

3 OF 4

ED

060 686


1.0
4.5
5.0
5.6
6.3
7.1
8.0
9.0
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

The Temporal Realization of Morphological and Syntactic Boundaries

Ilse Lehiste

Abstract

This paper is concerned with the effect of morphological and syntactic boundaries on the temporal structure of spoken utterances. Two speakers produced twenty tokens each of four sets of words consisting of a monosyllabic base form, disyllabic and trisyllabic words derived from the base by the addition of suffixes, and three short sentences in which the base form was followed by a syntactic boundary, this in turn followed by a stressed syllable, one unstressed syllable, and two unstressed syllables. The sentences thus reproduced the syllabic sequences of the derived words. The duration of words and segments was measured from oscillograms. The manifestation of morphological and syntactic boundaries is discussed, and some implications of the findings relative to the temporal programming of spoken utterances are considered.

0. Introduction

This paper is concerned with the effect of morphological and syntactic boundaries on the temporal structure of spoken utterances. The investigation was prompted by the observation made in the course of a previous study, [1,2] that the duration of a word may be considerably reduced, if a derivational suffix is added to the word constituting the base. In this earlier study, the words stead, skid and skit were compared with steady, skiddy and skitty. It might have been expected that the latter set would be longer than the former by the average duration of the derivational suffix. It turned out instead that the duration of the base part of the derived word was considerably shortened, so that even with the addition of a fairly long -y, the overall duration of the derived words was not much different from that of the base words.

In the current study, four sets of words were examined, built around the base forms stick, sleep, shade, and speed. Each of the words occurred by itself and in eight additional utterance types. Five derivational suffixes were used, three of them monosyllabic and two disyllabic. The words were further placed in short sentences in which they were followed by a major syntactic boundary--the boundary between the noun phrase functioning as subject and the verb phrase functioning as predicate. The verb phrase itself either consisted of a stressed monosyllable (in three cases) or started with a stressed syllable (in one case); or it started with one or two unstressed syllables. The sentences thus reproduced the syllabic sequences of the derived words. It was the purpose of the study to explore whether there are any differences in the durations of the base, depending on whether it is followed by a morpheme boundary within the same word, or by a major syntactic boundary coinciding with the word boundary.

I. Method

The test material, presented in Table 1, was recorded by two speakers, R.G. (male) and L.S. (female), both graduate students at The Ohio State University. The recordings were made under standard conditions in an anechoic chamber using reliable recording equipment. The utterances were produced in two ways, to test the comparability of different contexts and to vary the fairly artificial recording technique of repeating the same word a large number of times. One of the ways was indeed the repetition technique: each word was uttered ten times under a subjectively established 'constant' rate. Then each set, consisting of base word, derived words, and three short sentences, was read ten times in succession. Each speaker thus produced 20 tokens of each word, for a total of 720 utterances by each speaker.

The durations of words and segments were measured from oscillograms, produced by processing the recorded tapes through a Frøkjær-Jensen Trans-Pitch Meter and Intensity Meter, connected to a four-channel Elema-Schönander Mingograph. The material was analyzed statistically, using the IBM 360 Model 75 computer available at The Ohio State University Instruction and Research Computer Center.

II. Comparability of the Two Sets of Data

For both sets of data, the following computations were carried through: the mean duration of each segment; the mean duration of each word; the mean duration of the base component of the derived word (e.g., stick in sticky); the variances and standard deviations of each segment and word. The differences between the corresponding means for each segment and word were tested for significance according to the formula:

$$(1) \qquad Z = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{\sigma^2_A}{N_A} + \frac{\sigma^2_B}{N_B}}}$$

The difference in variability between the two sets was tested by two (related) measures:[3]

$$(2) \qquad H = \frac{\sigma^2_{MAX}}{\sigma^2_{MIN}} \qquad\qquad C = \frac{\sigma^2_{MAX}}{\sigma^2_{MIN} + \sigma^2_{MAX}}$$

For the given number of tokens, the critical values (at the 95% confidence level) were 1.960 for Z, 4.030 for H, and 0.801 for C.

It was found that the differences between the two sets of utterances for each speaker were random, and that there was minimal overlap between the two speakers in cases of statistically significant differences. Out of 196 pairwise comparisons of $\bar{X}_A$ and $\bar{X}_B$, speaker R.G. had 65 significant differences, speaker L.S. 88 significant differences; the same segments were involved in 35 instances, but these segments constituted no natural set: there was no discernible system. A separate check of syllable nuclei showed 11 instances for R.G. and 26 instances for L.S. in which the means differed significantly, i.e. Z was higher than the critical value. The same syllable nucleus was involved in 9 instances. As regards the differences in variability between the two sets, speaker R.G. had 15 (out of 196) cases in which the difference in variances between the two sets was significant; speaker L.S. had 36 instances, of which 9 involved the same segment for both speakers. As far as syllable nuclei were concerned, speaker R.G. had 2 instances of significant differences, L.S. 4, with an overlap of 2.

Combining the two sets would tend to increase the extreme ranges for each combined set of utterances and thus increase the variability; but since the difference in variability between the two sets was negligible, it was decided to combine the two sets in future calculations. The resultant increase in variability was in effect quite small. It is hoped that the method of producing the test utterances in the two different ways described above will have reduced the artificiality of the situation in which long lists of words are produced out of context, and that the results are better applicable to a more natural speech situation.

III. Effect of Morpheme Boundaries

In order to study the effect of morpheme boundaries (and word boundaries) on the duration of the base to which derivative suffixes were added, B/D ratios were computed. This term refers to the ratio of the durations of the base word (produced by itself) and the sum of the durations of the same segments occurring in the derived word (e.g., the mean duration of stick would be divided by the mean duration of the stick part of the word sticky). These ratios were calculated for all test words,and, separately, for the syllable nuclei in all test words. The differences between the means were highly significant in all instances; Z-values, which were always higher than the critical value, will not be included in the tables. The results are presented in Tables II-V and graphically in Figures 1 - 4. The tables are self-explanatory; a few words of explanation may be needed for the figures.

On each figure, the derived word types and sentence types are given on the vertical axis. The horizontal axis is calibrated to show increasing B/D ratios. Points representing B/D ratios for words are connected with solid lines; points representing B/D ratios for syllable nuclei are connected with dashed lines. The curves start in the left hand top corner at the B/D value 1: Base/Base yields a ratio of 1. Increasing ratios show decrease in the duration of the base component of the derived word resp. its syllable nucleus.

Several observations may be made regarding the figures. In no case was the duration of the same set of segments greater in a derived word than in the base form. The suffixes -*y*, -*er* and -*ing* seem to be equivalent with respect to their effect on the duration of the stem. It appears that the number of segments in the suffix has no systematic effect on the duration of the stem. This observation is confirmed by looking at the behavior of stem forms before the suffix -*ily*. This suffix was in fact pronounced with a syllabic /l/ by both speakers in all productions; thus the stems of words like *sticking* and *stickily* were followed by two segments each, but the -*ing* suffix was monosyllabic and the -*ily* suffix was disyllabic. In all cases, the disyllabic -*ily* suffix produced greater reduction in the duration of the stem than the monosyllabic suffix -*ing*, although both consisted of the same number of segments.

The suffix -*iness* constitutes a special case. In each instance, the B/D ratio was greatest under this condition. This is a disyllabic suffix, as is -*ily*; however, its rhythmic structure is considerably different. It seems possible that in the case of the -*iness* suffix we are dealing with two cycles of derivation: that, for example, *sticky* is derived from *stick* in the first cycle, and *stickiness* from *sticky* in the second cycle. If this is so, then the ratios of *stick /sticky* and *sticky/stickiness* (involving the base forms *stick* and *sticky* respectively) should be approximately equal. Some support for this assumption may indeed be found in Table VI, which presents the pertinent ratios.

A comparison of the curves for words with the curves for syllable nuclei indicates that the reduction in the duration of a stem in the derived form is achieved more at the expense of vowels than at the expense of consonants. The nature of the vowel and the postvocalic consonant seem to play an equally important role. Intrinsically long syllable nuclei (like those in *sleep*, *speed*, and *shade*) are more compressible than intrinsically short syllable nuclei (as in *stick*). But /i/ in *sleep*, when followed by a voiceless plosive, is much less compressible than /i/ in *speed* and /e^{I}/ in *shade*. Tendencies for being reduced under a certain condition become accentuated when one looks at the most compressible segment: for both speakers, the greatest effects of the various positions are manifested in the syllable nuclei of *speed* and *shade*.

IV. Effect of Syntactic Boundaries

One of the hypotheses tested in this experiment was the hypothesis that syntactic boundaries would have temporal effects that are clearly distinct from those of morpheme boundaries. However, the results of this study show that as far as the temporal structure of utterances is concerned, effects of morpheme boundaries and effects of syntactic boundaries cannot be separated from each other. Furthermore, it is not certain that the boundaries as such have any effect at all, since the temporal structure of the utterances seems to depend most of all on their syllabic structure, regardless of the nature of the boundaries involved.

In sentences like *Speed kills*, we find durations of the test word that are very similar to those of disyllabic bimorphemic words;

sentences like The speed increased resemble most words like speediness, with an unstressed short syllable followed by a relatively long syllable. The addition of another unstressed syllable may have a further reducing effect, but the data are not consistent at this point. The major result here is the absence of any clear differences between the effects of morpheme boundaries and syntactic boundaries, and the likelihood that the durational structure is conditioned by the number of syllables rather than either by the number of segments or by the presence of boundaries.

V. Generality of the Findings

One of the ways to test the results would be to form predictions on the basis of these data and then compare the predictions with further observations. I intend to record other sets of words by the same speakers as well as the same sets of words by different speakers, and calculate the goodness of fit between predicted and observed B/D ratios. The basis for predictions might be Table VII, which combines words that seem to behave in a similar fashion for the two speakers.

VI. Discussion

The results of this study confirm earlier studies in some respects, but differ from them in certain important aspects.

Bolinger[4] stated that long syllables tend to acquire extra length if followed by another long syllable (long syllables being those that contain a full vowel); if followed by a short syllable, long syllables cannot acquire that extra length and therefore appear shorter. This process tends to ignore morpheme and word boundaries, and may take place across a syntactic boundary.

The present study confirms Bolinger's notion that temporal readjustment processes tend to ignore morpheme and word boundaries. The shortening of a long syllable before a short syllable is likewise confirmed in all the data. However, in sentences of the type Speed kills, the word speed (and words in analogous sentences) certainly did not acquire any extra length, at least in comparison to isolated productions of the same word.

Gaitenby[5] found a common ratio of segment-to-utterance length for all dialects of American English sampled in her study. When segment durations were converted to percentages of total utterance time, it was found that 90% of all the segments varied less than 5.3% for any speaker. The longer the utterance in terms of number of segments, the shorter the absolute duration of any given segment, until an approximate minimum duration was reached beyond which segments could not be compressed any further. She noted also that words immediately preceding a pause tended to expand in utterances of all lengths. According to Gaitenby, it would thus be the word closest to the pause that would acquire extra length, while in longer utterances, the preceding parts of the sentence would be produced at a faster rate. This seems to be borne out by the findings: in the three sentences, the base word became successively shorter, the farther it was removed from the end of the sentence. A difference between Gaitenby's results and those obtained in this study is the observation that utterance length should be determined with reference to number and type of syllables rather than with reference to the number of segments.

Chomsky and Halle[6] have postulated a hierarchy of boundaries which delimit linguistic units that serve as domains of application of different kinds of phonological rules. Although the authors are careful to state that phonetic effects need not be associated with (word) boundaries, the postulation of a hierarchy of boundaries naturally prompts a phonetician to look for possibly hierarchical differences in the manifestations of these boundaries. I had previously formulated the hypothesis that phonological units are definable in terms of suprasegmental patterns, while their boundaries are mainly manifested in terms of modifications of segments.[7] Few, if any, indications of word boundaries emerged from the present study. There were a small number of instances in which the duration of the segment preceding a word boundary was greater than the duration of the same segment preceding a morpheme boundary. As far as the overall temporal organization of the utterances is concerned, no evidence for a hierarchical organization of boundaries was found as a result of this study. The temporal organization of spoken language seems to take place in terms of speech production units which are fairly independent of the morphological or syntactic structure of the utterances.

Acknowledgements

Grateful acknowledgement is made of the help of Mr. Thomas G. Whitney of the Ohio State University Instruction and Research Computer Center, who wrote the computer programs employed in this study.

Footnotes

[1] I. Lehiste, "The Temporal Organization of Higher-Level Linguistic Units," Paper presented at the April 1970 meeting of the Acoustical Society of America, Atlantic City, N.J. (1970).

[2] I. Lehiste, "Temporal Organization of Spoken Language," In: Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen, Ed. by L. L. Hammerich, Roman Jakobson, and Eberhard Zwirner (Akademisk Forlag, Copenhagen, 1971), pp. 275-285.

[3] B. J. Winer, Statistical Principles in Experimental Design (McGraw-Hill, New York, 1962), p. 94.

[4] D. Bolinger, "Length, Vowel, Juncture," Linguistics 1.1.5-29 (1963). (To be revised).

[5] J. Gaitenby, "The Elastic Word," Paper given at the Tenth Annual National Conference on Linguistics, sponsored by the Linguistic Circle of New York, 13 March 1965. Status Report on Speech Research SR-2 (Haskins Laboratories, New York, 1965), pp. 3.1-3.12.

[6] N. Chomsky and M. Halle, The Sound Pattern of English (Harper & Row, New York, 1968).

[7] I. Lehiste, Suprasegmentals (M.I.T. Press, Cambridge, Mass., 1970).

Table I. Test materials used in the study. The symbol - is used to indicate the boundary between stem and derivative suffix. # symbolizes word boundary; $\acute{\_}$ and $\breve{\_}$ refer to stressed and unstressed syllables.

BASE	stick	sleep	shade	speed
-Y	sticky	sleepy	shady	speedy
-ER	sticker	sleeper	shader	speeder
-ING	sticking	sleeping	shading	speeding
-ILY	stickily	sleepily	shadily	speedily
-INESS	stickiness	sleepiness	shadiness	speediness
# $\acute{\_}$	the stick fell	sleep heals	the shade lingered	speed kills
# $\breve{\_}\,\acute{\_}$	the stick is broken	sleep refreshes	the shade increased	the speed increased
# $\breve{\_}\,\breve{\_}\,\acute{\_}$	the stick was discarded	my sleep was disturbed	the shade was refreshing	the speed was controlled

Table II. Mean durations (in milliseconds), standard deviations and B/D ratios for two sets of words and corresponding syllable nuclei produced by speaker R.G.

Utterance	Duration of base	σ	B/D ratio	Duration of Syl. nucleus	σ	B/D ratio
stick	401.55	29.45		130.70	6.94	
sticky	312.80	23.68	1.284	93.45	6.53	1.399
sticker	302.50	17.49	1.327	89.45	8.85	1.461
sticking	295.45	16.92	1.359	88.80	7.28	1.472
stickily	291.10	17.90	1.379	84.15	6.75	1.553
stickiness	265.75	15.79	1.511	78.90	5.63	1.657
The stick fell	274.85	14.10	1.461	87.90	7.02	1.487
The stick is broken	248.20	12.65	1.618	81.65	7.57	1.601
The stick was discarded	245.10	13.49	1.638	77.90	5.81	1.678
sleep	409.80	18.96		123.55	14.55	
sleepy	336.80	19.70	1.217	84.15	7.97	1.468
sleeper	341.25	19.83	1.201	83.10	9.21	1.487
sleeping	330.35	18.12	1.241	81.50	10.11	1.516
sleepily	313.35	13.99	1.308	69.60	8.58	1.775
sleepiness	287.05	13.81	1.428	62.05	6.79	1.991
sleep heals	305.95	16.33	1.339	75.95	8.41	1.627
sleep refreshes	299.60	19.90	1.368	61.85	4.67	1.998
My sleep was disturbed	307.45	17.44	1.333	59.65	9.65	2.071

Table III. Mean durations (in milliseconds), standard deviations and B/D ratios for two sets of words and corresponding syllable nuclei produced by speaker L.S.

Utterance	Duration of base	σ	B/D ratio	Duration of Syl. nucleus	σ	B/D ratio
stick	431.80	43.33		168.90	23.25	
sticky	346.00	34.44	1.248	115.50	15.83	1.462
sticker	331.95	25.88	1.301	109.65	14.75	1.540
sticking	348.30	30.56	1.240	109.20	17.36	1.547
stickily	303.10	17.93	1.425	77.05	6.89	2.192
stickiness	271.60	20.78	1.590	76.50	6.92	2.208
The stick fell	311.15	22.74	1.388	91.35	11.17	1.849
The stick is broken	283.90	19.46	1.521	88.85	10.83	1.901
The stick was discarded	268.15	28.40	1.610	80.75	8.42	2.092
sleep	442.45	39.62		180.30	16.85	
sleepy	363.40	19.64	1.218	131.45	9.24	1.372
sleeper	363.35	22.87	1.218	127.25	8.90	1.417
sleeping	374.45	18.26	1.182	132.45	10.87	1.361
sleepily	342.60	16.72	1.291	114.50	8.72	1.575
sleepiness	307.70	16.39	1.438	96.55	8.45	1.867
Sleep heals	325.00	25.33	1.361	113.55	14.77	1.588
Sleep refreshes	282.75	18.96	1.565	93.55	9.74	1.927
My sleep was disturbed	314.90	26.82	1.405	99.40	19.27	1.814

Table IV. Mean durations (in milliseconds), standard deviations and B/D ratios for two sets of words and corresponding syllable nuclei produced by speaker R.G.

Utterance	Duration of base	σ	B/D ratio	Duration of Syl. nucleus	σ	B/D ratio
speed	511.50	34.95		266.00	28.17	
speedy	359.75	15.09	1.422	150.50	10.25	1.767
speeder	344.75	16.42	1.484	141.50	11.01	1.880
speeding	342.50	13.13	1.493	136.00	9.81	1.956
speedily	322.50	18.03	1.586	120.00	8.27	2.217
speediness	313.25	16.57	1.633	115.50	7.76	2.303
Speed kills	344.00	17.06	1.487	125.50	8.87	2.120
The speed increased	301.25	15.12	1.698	110.00	7.61	2.418
The speed was controlled	293.50	20.53	1.743	104.00	8.97	2.558
shade	454.10	28.88		266.15	18.61	
shady	327.20	20.08	1.388	181.85	14.79	1.464
shader	324.20	18.81	1.401	172.40	9.54	1.544
shading	306.95	23.39	1.479	158.00	11.24	1.684
shadily	276.70	10.20	1.641	132.05	8.74	2.016
shadiness	265.20	17.60	1.712	125.35	9.83	2.123
The shade lingered	324.80	18.49	1.398	146.95	16.23	1.811
The shade increased	298.60	18.44	1.521	130.15	12.93	2.045
The shade was refreshing	307.60	26.05	1.476	131.50	18.61	2.024

Table V. Mean durations (in milliseconds), standard deviations and B/D ratios for two sets of words and corresponding syllable nuclei produced by speaker L.S.

Utterance	Duration of base	σ	B/D ratio	Duration of Syl. nucleus	σ	B/D ratio
speed	574.25	30.00		297.85	16.25	
speedy	394.85	23.89	1.454	163.30	11.69	1.824
speeder	403.85	18.44	1.422	171.75	13.52	1.734
speeding	396.10	24.54	1.450	158.75	12.86	1.876
speedily	354.50	29.75	1.620	126.25	16.98	2.359
speediness	322.70	23.41	1.780	104.40	6.66	2.853
Speed kills	416.55	27.28	1.379	163.05	19.07	1.827
The speed increased	342.85	20.97	1.675	127.30	11.68	2.340
The speed was controlled	305.50	22.00	1.880	96.65	7.92	3.082
shade	454.65			267.70	22.88	
shady	321.65	20.72	1.413	165.25	11.26	1.620
shader	326.75	26.61	1.391	160.50	14.16	1.668
shading	312.95	22.09	1.453	159.30	19.72	1.680
shadily	294.15	26.41	1.546	139.95	21.89	1.913
shadiness	261.65	25.37	1.738	112.55	11.65	2.378
The shade lingered	331.95	36.24	1.370	154.40	23.93	1.734
The shade increased	282.20	22.03	1.611	135.75	16.90	1.972
The shade was refreshing	273.40	23.97	1.633	114.25	15.97	2.343

Table VI. Mean durations (in milliseconds), standard deviations, and B/D ratios for words derived with -y and -ness, in which the -ness words are derived by a two-cycle operation from the base.

Speaker R.G.				Speaker L.S.			
Utterance	Duration of base	σ	B/D ratio	Utterance	Duration of base	σ	B/D ratio
stick	401.55	29.45		stick	431.80	43.33	
stick-y	312.80	23.68	1.284	stick-y	346.00	34.44	1.248
sticky	513.25	37.52		sticky	557.45	36.59	
sticky-ness	376.75	17.66	1.362	sticky-ness	388.50	24.34	1.435
sleep	409.70	18.96		sleep	442.45	39.62	
sleep-y	336.70	19.70	1.217	sleep-y	363.40	19.64	1.218
sleepy	517.55	26.58		sleepy	544.20	30.99	
sleepy-ness	369.65	14.15	1.400	sleepy-ness	392.20	18.46	1.388
speed	511.50	34.95		speed	574.25	30.00	
speed-y	359.75	15.09	1.422	speedy	394.85	23.89	1.454
speedy	529.95	26.23		speedy	597.40	16.94	
speedy-ness	396.35	16.60	1.337	speedy-ness	410.55	31.19	1.455
shade	454.10	28.88		shade	454.65	35.84	
shade-y	327.20	20.08	1.388	shade-y	321.65	20.72	1.413
shady	477.90	25.81		shady	490.60	24.43	
shady-ness	346.30	16.46	1.380	shady-ness	329.70	23.87	1.488

Table VII. Average B/D ratios (speakers R.G. and L.S. combined)

	stick, sleep		shade, speed	
	WORD	SN	WORD	SN
Base	1.00	1.00	1.00	1.00
-Y	1.242	1.425	1.420	1.669
-ER	1.262	1.476	1.425	1.706
-ING	1.256	1.474	1.469	1.799
-ILY	1.351	1.774	1.599	2.126
-INESS	1.492	1.931	1.716	2.415
# $\underline{\acute{\ }}$	1.388	1.638	1.409	1.873
# $\underline{\breve{\ }}\ \underline{\acute{\ }}$	1.518	1.857	1.626	2.194
# $\underline{\breve{\ }}\ \underline{\breve{\ }}\ \underline{\acute{\ }}$	1.497	1.914	1.683	2.502

Fig. 1. B/D ratios for the words stick and sleep and their syllable nuclei for speaker LS. The base word and the derivative forms are indicated on the vertical axis; the horizontal axis is calibrated for ratios of duration of base word/duration of the base part of the derived word.

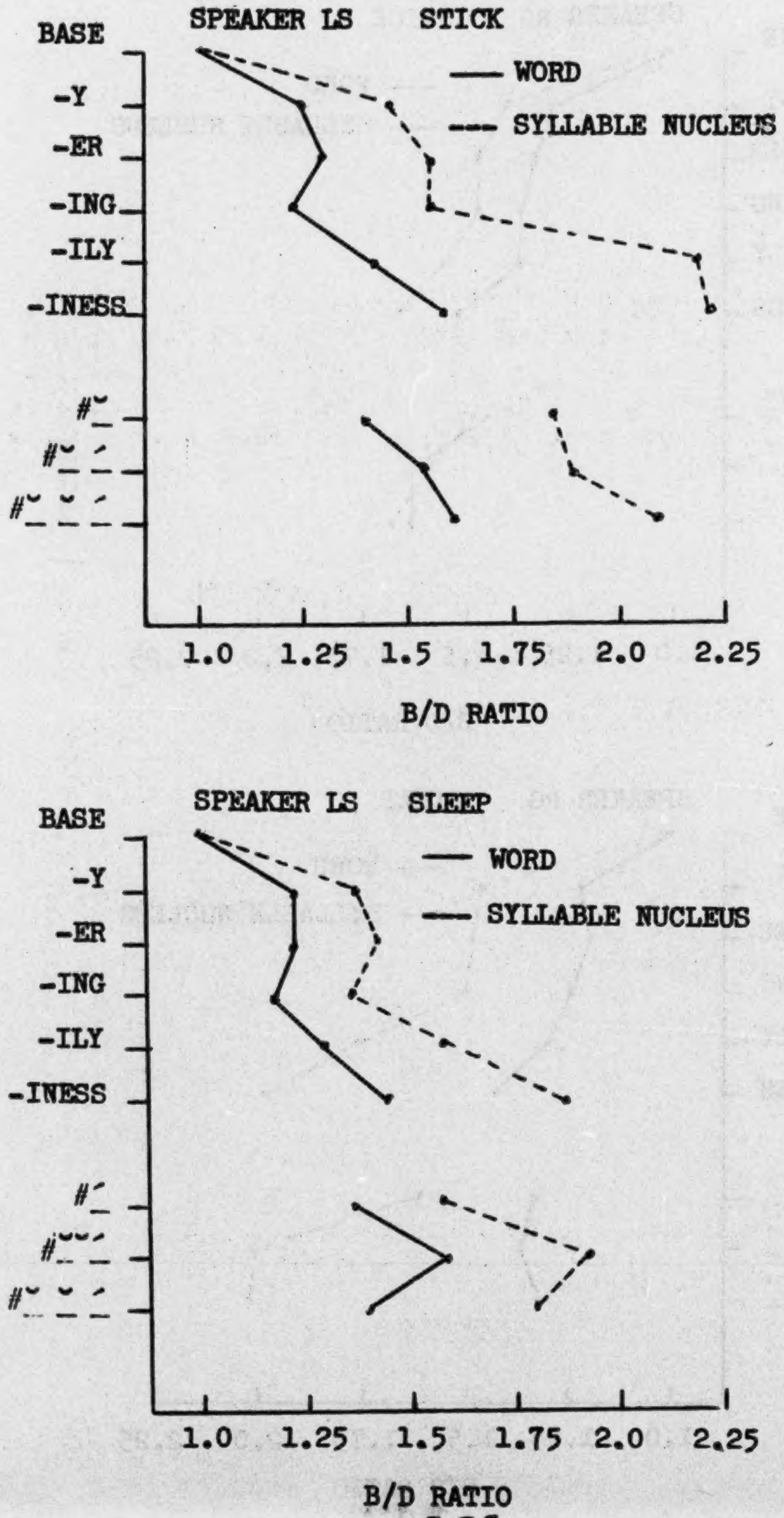


Fig. 2. B/D ratios for the words stick and sleep and their syllable nuclei for speaker RG. The base word and the derivative forms are indicated on the vertical axis; the horizontal axis is calibrated for ratios of duration of base word/duration of the base part of the derived word.

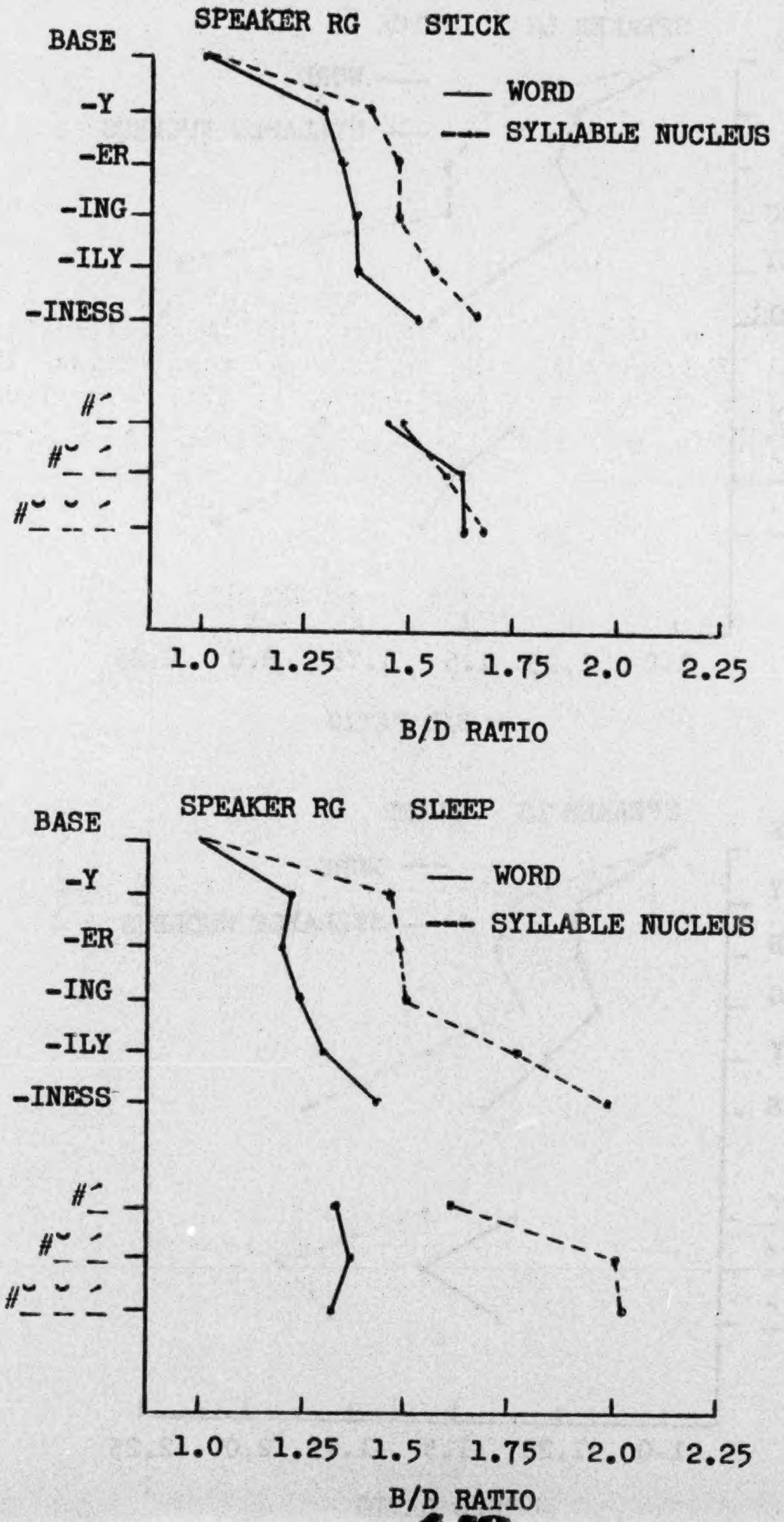


Fig. 3. B/D ratios for the words speed and shade and their syllable nuclei for speaker RG. The base word and the derivative forms are indicated in the vertical axis; the horizontal axis is calibrated for ratios of duration of base word/duration of the base part of the derived word.

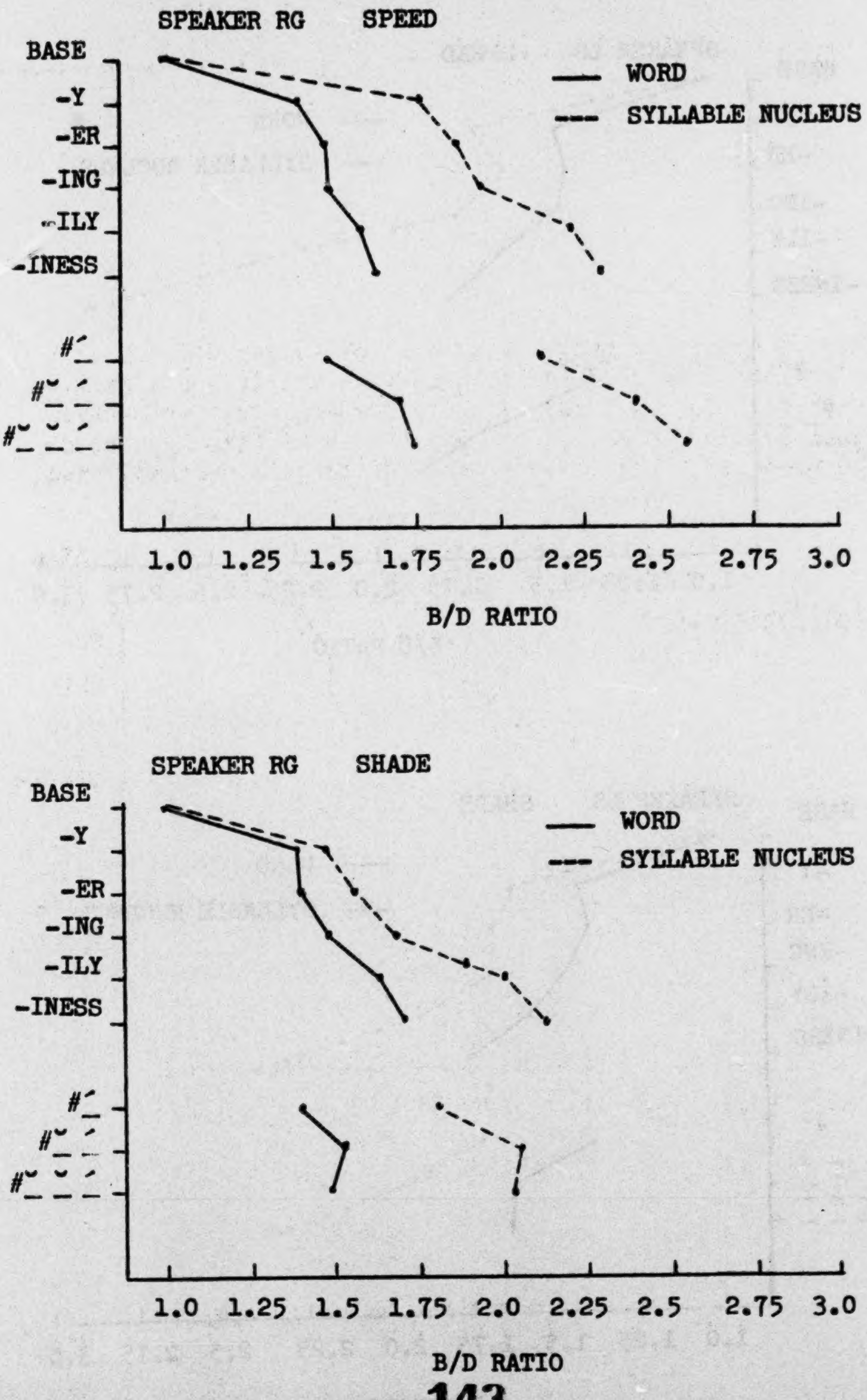
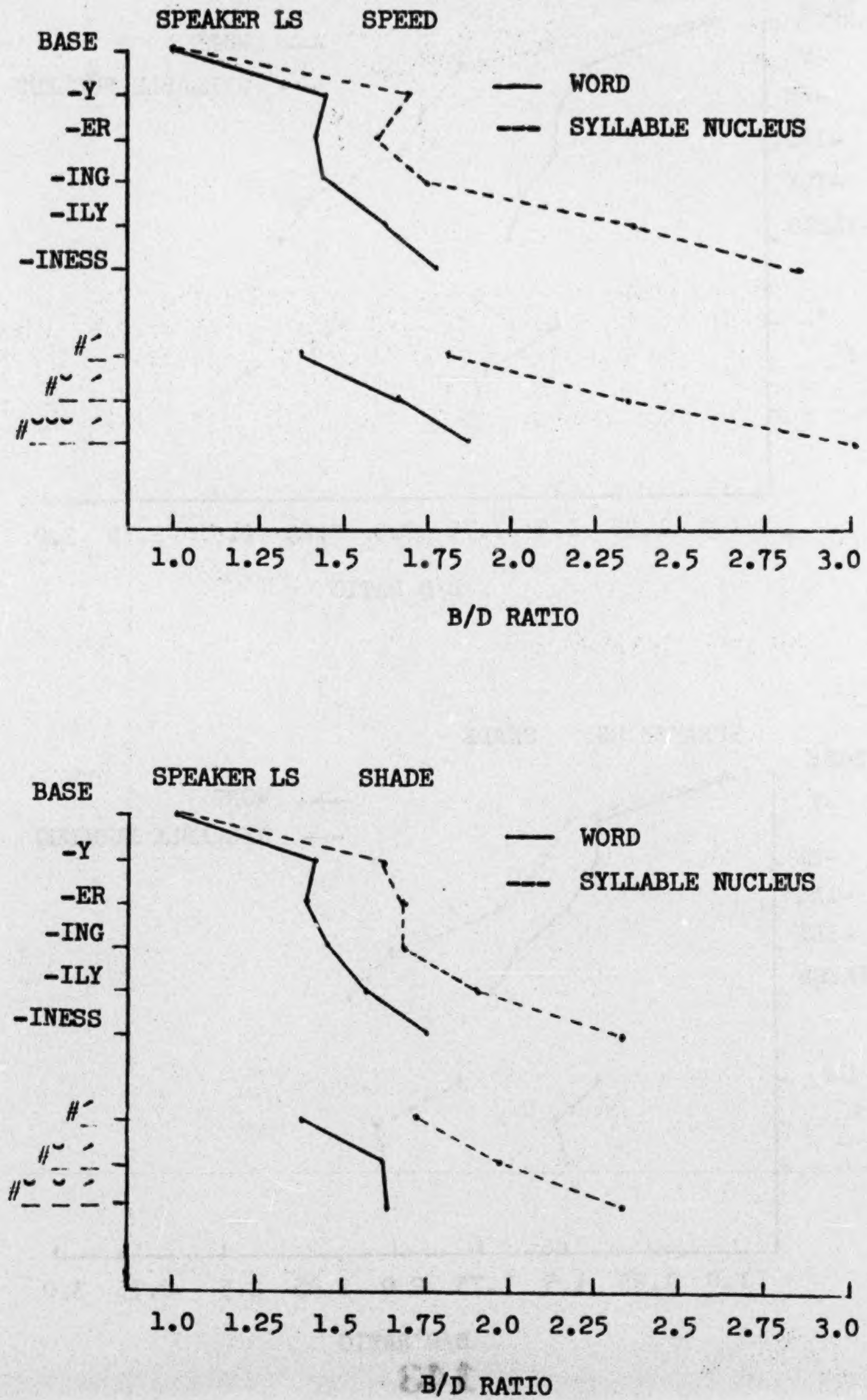


Fig. 4. B/D ratios for thw words speed and shade and their syllable nuclei for speaker LS. The base word and the derivative forms are indicated on the vertical axis; the horizontal axis calibrated for ratios of duration of base word/duration of the base part of the derived word.



Comparison of Controlled and Uncontrolled Normal Speech Rate*

Richard Gregorski, Linda Shockey,

and Ilse Lehiste

---

*Sponsored in part by the National Science Foundation through Grant GN 534.1 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

Comparison of Controlled and Uncontrolled Normal Speech Rate

Richard Gregorski, Linda Shockey, and Ilse Lehiste

Temporal studies have employed basically two methods for elicitation of speech rate: 1) controlled, i.e., externally induced through the use of a pulsating beat, and 2) uncontrolled, i.e., internally generated by the subject with the instruction to maintain a constant rate. Peterson and Lehiste (1960) in investigating the influence of tempo on the duration of syllable nuclei had their subjects "speak in synchronism with a periodic pulse." Lindblom (1963) used periodic clicks to manipulate speech rate in examining vowel reduction under varying tempos. Kozhevnikov and Chistovich (1965) in their experiment on the effect of rate on relative speech durations employed as a rate control a low-frequency periodic oscillation generator which was triggered by the subject's initiation of articulation. However, in their experiment to determine the number of articulatory programs in a sentence of two syntagmas, no external device was used to control rate; instead, the speaker was "instructed to adhere during all pronunciations to one and the same rate of speech." In their experimental check of syllable command hypotheses using multiple repetitions of a sentence, the subjects performed the task first at a rapid rate and then at a slow rate; no external control appears to have been employed. Nooteboom and Slis (1969) in their speech rate study had their subjects freely choose their fast, normal, and slow rates. Lehiste (1970b) in her study of the temporal organization of monosyllabic and disyllabic words in English had her subjects maintain a "subjectively constant rate."

To our knowledge, the comparability of the durations of speech units produced at a subjectively determined rate and those produced at a rate controlled by an external source has never been determined. If significant differences exist between temporal patterns occurring in speech produced by the two methods of elicitation, obvious questions arise. For example, to what extent could we then generalize about the temporal organization of speech from the previously mentioned studies executed with non-comparable methods? Would not the differences perhaps suggest two types of programming: 1) a basic language program including speech-unit organization and natural rhythm information, and 2) a synchronization program whose task is to adjust the language program until its natural rhythm is synchronous with the external rhythm?

It was the purpose of this experiment to determine the comparability of controlled and uncontrolled normal speech rate for both a sentence and a word spoken in isolation. Aggie was chosen for the word, and I bag Aggie, for the sentence. The major criterion for selecting these utterances was their relatively segmentable structure when converted into oscillographic displays, and not their high semantic content. Two native speakers of English were instructed to produce both the word and the sentence about 150 times each at a comfortably constant normal rate. From recordings of these productions oscillograms were made by use of a Frøkjaer-Jensen transpitch meter and an Elema-Schönander Mingograph (100 mm/sec). Durations of individual segments and pauses were measured to the nearest 1/2 millimeter (i.e., 5 milliseconds). The mean duration, standard deviation, variance, and coefficient of variation ($\frac{\sigma}{M}$) were computed using an IBM 360 computer for all possible combinations of adjacent segments.

A Seth Thomas electronic metronome was used to implement the control method. To obtain the pulse rate for the controlled utterances, the mean duration for each speaker's interstress interval for both the word and the sentence of the uncontrolled productions was converted into an equivalent pulse interval on the metronome. Since for both speakers the natural sentence stress fell on the /æ/ of *Aggie*, it was decided to synchronize the click with this stress. The speakers were instructed to repeat the production task, only this time synchronizing the /æ/ of *Aggie* with the click of the metronome. The same segmentation procedures and statistical analyses that were used for the uncontrolled utterances were applied to the controlled ones. The differences between the coefficients of variation of the controlled and uncontrolled sets were computed (see Tables I-VI in the Appendix).

Figure I presents the coefficient of variation comparisons of Speaker PM's controlled and uncontrolled *Aggie* spoken in isolation. There was an average difference of 2% in the coefficients of variation for segments. Notice that there was no difference between the coefficients of variation of the stressed /æ/'s; in absolute terms there was only a 10 millisecond difference in their mean durations. The syllables, word and word + pause likewise had average coefficient differences of about 2%. There was a 6% difference for the pauses.

Figure II presents the coefficient of variation comparisons of Speaker LS's controlled and uncontrolled *Aggie*. Her average coefficient differences for both segments and syllables were about 1 1/2%. There was a .3% difference for the word.

Figures III and IV present the coefficient comparisons for Speaker PM's controlled versus uncontrolled sentences. Segments, syllables, and words as groups had average coefficient differences of 1-2%. There was a 1% difference for the sentence and a .1% difference for the sentence + pause.

Figures V and VI present Speaker LS's sentence comparisons. Segments, syllables, and words as groups had average coefficient differences of 1-2%. There was a 1% difference for the sentence and a 3% difference for the sentence + pause.

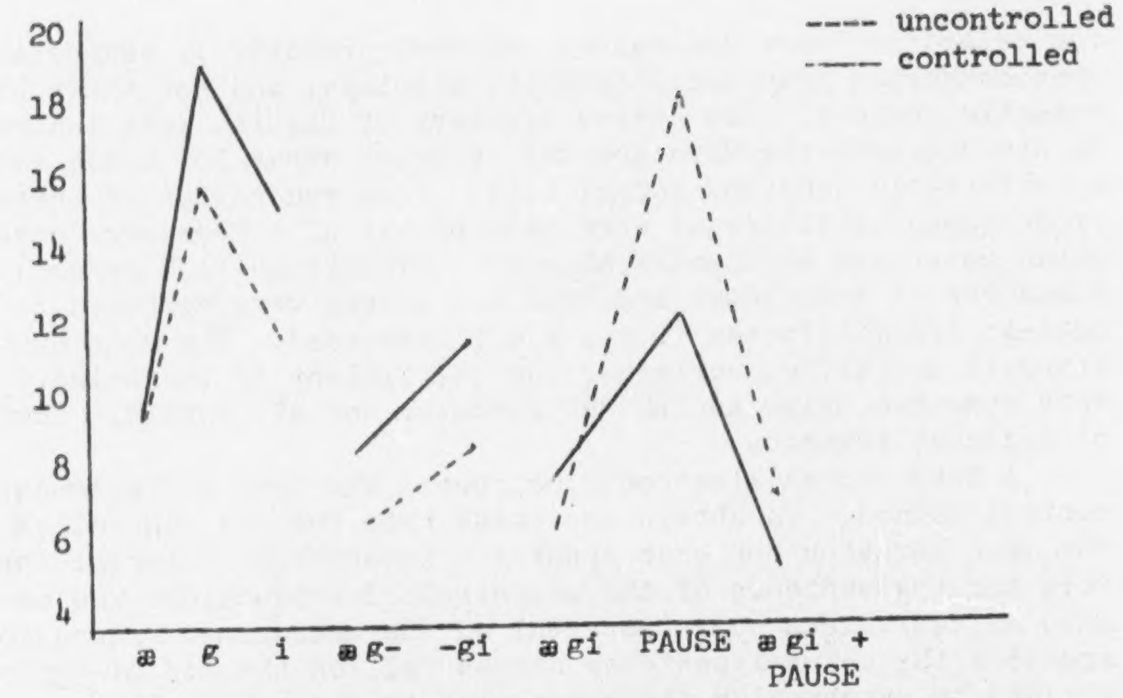


Figure I. Coefficient of variation ($\frac{\sigma}{M}$ x 100) comparisons of controlled versus uncontrolled speech-units for *Aggie* produced by speaker PM.

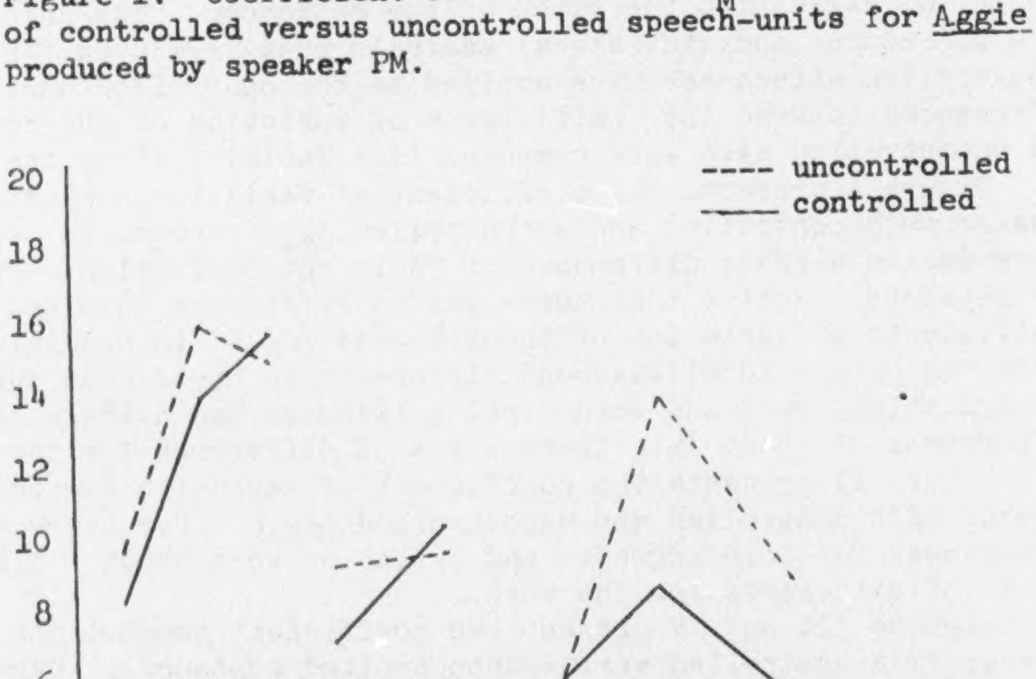


Figure II. Coefficient of variation ($\frac{\sigma}{M}$ x 100) comparisons of controlled versus uncontrolled speech-units for *Aggie* produced by Speaker L.S.

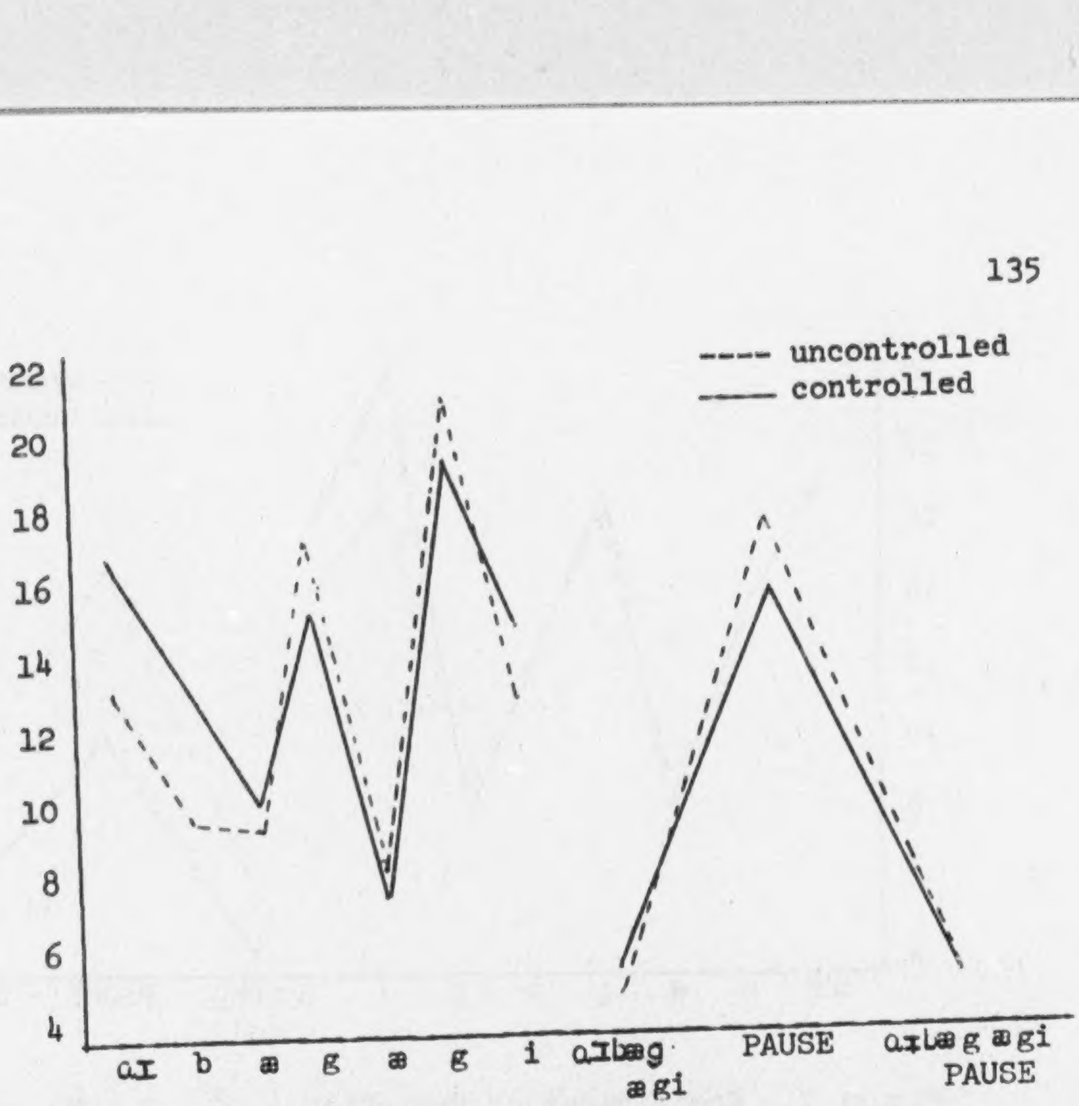


Figure III. Coefficient of Variation ($\frac{\sigma}{M}$ x 100)comparisons of controlled versus uncontrolled speech-units for *I bag Aggie* produced by speaker PM.

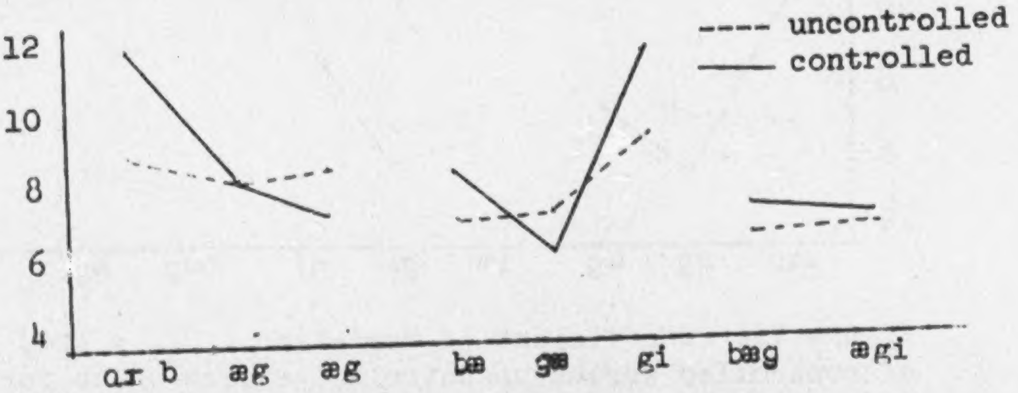


Figure IV. Coefficient of variation ($\frac{\sigma}{M}$ x 100) comparisons of controlled versus uncontrolled speech-units for *I bag Aggie* produced by speaker PM.

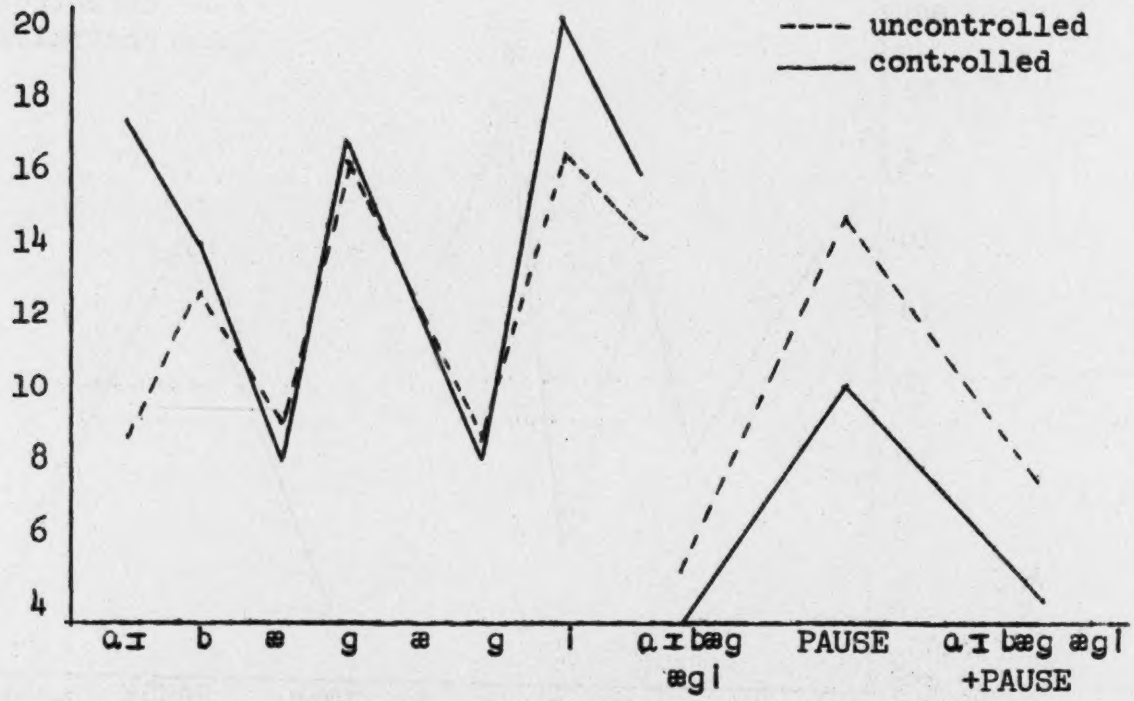


Figure V. Coefficient of variation ($\frac{\sigma}{M}$ x 100) comparisons of controlled versus uncontrolled speech-units for I bag Aggie produced by speaker LS.

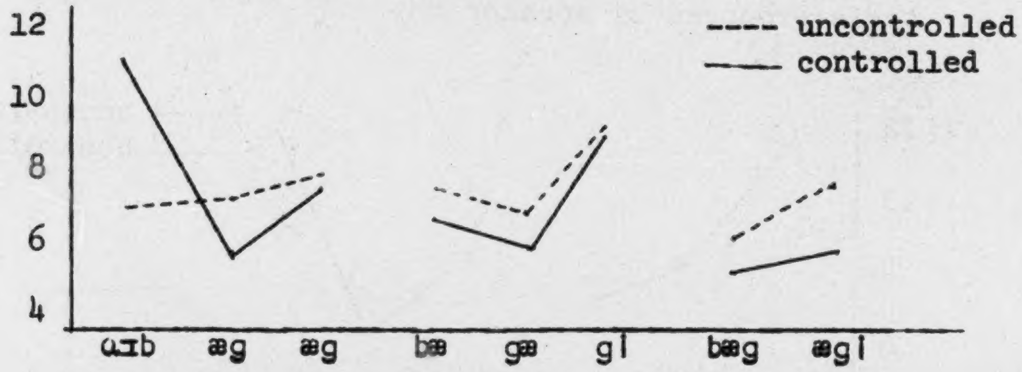


Figure VI. Coefficient of variation ($\frac{\sigma}{M}$ x 100) comparisons of controlled versus uncontrolled speech-units for I bag Aggie produced by speaker LS.

To test for the significance of these coefficient of variation differences, we assumed that if the same magnitude of difference exists between two uncontrolled sets and also between two controlled sets, then such differences cannot be attributed to the control technique. We divided both the controlled and uncontrolled sets into sequential halves of about 75 tokens each. The average coefficient differences between the uncontrolled halves and also between the controlled halves were comparable to those between the entire controlled and uncontrolled sets (see Table VII in the Appendix). It thus appears that these differences are due to the natural variability of speech in a repetition task and cannot be attributed to the use of the periodic beat.

The controlled and uncontrolled sets were also examined for the direction of the differences between the coefficients of variation. We found no systematic direction to these differences for either speaker.

We conclude that in repetitions of the same words and sentences spoken at a normal rate, the two methods described here produce comparable results. However, we want to emphasize that we make no claim regarding differences between controlled and uncontrolled speech produced at other rates or using other elicitation techniques.

Appendix

TABLE I

Coefficient of variation ($\frac{\sigma}{M}$) comparisons of uncontrolled versus controlled speech-units for Aggie produced by Speaker PM.

Speech-unit	Uncontrolled	Controlled	Difference	Average Difference
æ	.093	.093	--	.022
g	.158	.190	.032	
i	.116	.150	.034	
æg	.063	.080	.017	.022
gɪ	.087	.114	.027	
ægɪ	.060	.076	.016	.016
PAUSE	.184	.121	.063	.063
ægɪ + PAUSE	.069	.050	.019	.019

TABLE II

Coefficient of variation ($\frac{\sigma}{M}$) comparisons of uncontrolled versus controlled speech-units for Aggie produced by speaker LS.

Speech-unit	Uncontrolled	Controlled	Difference	Average Difference
æ	.096	.083	.013	.013
g	.157	.136	.021	
i	.149	.155	.006	
æg	.089	.067	.022	.016
gi	.094	.103	.009	
ægi	.061	.064	.003	.003
PAUSE	.136	.086	.050	.050
aægi + PAUSE	.085	.053	.032	.032

TABLE III

Coefficient of variation ($\frac{\sigma}{M}$) comparisons of uncontrolled versus controlled speech-units for I bag Aggie produced by speaker PM.

Speech-unit	Uncontrolled	Controlled	Difference	Average Difference
ɑ ɪ	.130	.168	.038	.022
b	.096	.131	.035	
æ$_1$	.092	.101	.009	
g$_1$	.168	.147	.021	
æ$_2$	.080	.072	.008	
g$_2$	.211	.185	.026	
i$_2$	.126	.146	.020	
ɑɪb	.087	.119	.032	.015
bæ	.067	.083	.016	
æg$_1$	.081	.081	--	
gæ$_1$	.070	.061	.009	
æg$_2$	.085	.071	.014	
gi$_2$	.093	.114	.021	

TABLE IV

Coefficient of variation ($\frac{\sigma}{M}$) comparisons of uncontrolled versus controlled speech-units for I bag Aggie produced by speaker PM.

Speech-unit	Uncontrolled	Controlled	Difference	Average Difference
ɑɪbæ	.070	.094	.024	.011
bæg	.063	.072	.009	
ægæ	.054	.057	.003	
gæg	.077	.062	.015	
ægi	.066	.068	.002	
ɑɪbæg	.068	.083	.015	.007
bægæ	.045	.054	.009	
ægæg	.057	.055	.002	
gægi	.061	.061	—	
ɑɪbægæ	.051	.063	.012	.007
bægæg	.049	.052	.003	
ægægi	.049	.054	.005	
ɑɪbægæg	.054	.059	.005	.006
bægægi	.044	.051	.007	
ɑɪbægægi	.047	.056	.009	.009
PAUSE	.176	.153	.023	.023
ɑɪbægægi + PAUSE	.050	.051	.001	.001

TABLE V

Coefficient of variation ($\frac{\sigma}{M}$) comparisons of uncontrolled versus controlled speech-units for I bag Aggie produced by Speaker LS.

Speech-unit	Uncontrolled	Controlled	Difference	Average Difference
ɑɪ	.107	.175	.068	
b	.126	.139	.013	
æ$_1$	.091	.082	.009	
g$_1$	.158	.165	.007	.023
æ$_2$	.086	.081	.005	
g$_2$	.159	.198	.039	
i	.140	.157	.017	
ɑɪb	.068	.109	.041	
bæ	.072	.063	.009	
æg$_1$	.070	.057	.013	.013
gæ	.065	.056	.009	
æg$_2$	.076	.072	.004	
gi	.091	.089	.002	

TABLE VI

Coefficient of variation ($\frac{\sigma}{M}$) comparisons of uncontrolled versus controlled speech-units for I bag Aggie produced by speaker LS.

Speech-unit	Uncontrolled	Controlled	Difference	Average Difference
ɑɪbæ	.063	.064	.001	.007
bæg	.059	.052	.007	
ægæ	.056	.046	.010	
gæg	.060	.059	.001	
ægi	.071	.056	.015	
ɑɪbæg	.055	.056	.001	.006
bægæ	.048	.042	.006	
ægæg	.054	.047	.007	
gægi	.061	.050	.011	
ɑɪbægæ	.051	.045	.006	.009
bægæg	.048	.044	.004	
ægægi	.056	.040	.016	
ɑɪbægæg	.050	.043	.007	.009
bægægi	.049	.039	.010	
ɑɪbægægi	.052	.039	.013	.013
PAUSE	.145	.102	.043	.043
ɑɪbægægi + PAUSE	.074	.045	.029	.029

TABLE VII

Coefficient of variation ($\frac{\sigma}{M}$) differences between various set comparisons of speech-units for Aggie and I bag Aggie produced by speakers PM and LS.

	Set Comparison *	Segments	Syllables	Word(s)	Sentence
Speaker PM Aggie	UNCON/UNCON	.014	.003	.009	▨
	CONT/CONT	.020	.014	.013	
	UNCON/CONT	.022	.022	.016	
Speaker LS Aggie	UNCON/UNCON	.051	.060	.070	▨
	CONT/CONT	.019	.020	.007	
	UNCON/CONT	.013	.016	.003	
Speaker PM I bag Aggie	UNCON/UNCON	.022	.018	.011	.020
	CONT/CONT	.029	.020	.020	.009
	UNCON/CONT	.022	.015	.006	.009
Speaker LS I bag Aggie	UNCON/UNCON	.015	.003	.005	.006
	CONT/CONT	.016	.006	.005	.012
	UNCON/CONT	.022	.013	.011	.013

*UNCON= Uncontrolled, CONT= Controlled.

Bibliography

Kozhevnikov, V. A., and L. A. Chistovich. Speech: Articulation and Perception. Translated by J.P.R.S., Washington, D.C., No. JPRS 30, 543. Moscow-Leningrad. 1965.

Lehiste, Ilse. Suprasegmentals, Cambridge: M.I.T. Press, 1970a.

Lehiste, Ilse. "Temporal Organization of Spoken Language," Working Papers in Linguistics No. 4, 96-113. Ohio State University, Columbus, Ohio. 1970b.

Lindblom, B. "Spectrographic Study of Vowel Reduction," Journal of The Acoustical Society of America 35, 1773-1781. 1963.

Nooteboom, S. G., and I. H. Slis. "A Note on Rate of Speech," IPO Annual Progress Report, No. 4, 58-60. Institute for Perception Research, Eindhoven, Holland. 1969.

Peterson, G. and Ilse Lehiste. "Duration of Syllable Nuclei in English," Journal of the Acoustical Society of America 32, 693-703, 1960.

Word Unit Temporal Compensation*

Linda Shockey, Richard Gregorski,

and Ilse Lehiste

---

*Sponsored in part by the National Science Foundation through Grant GN-534.1 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

Word Unit Temporal Compensation

Linda Shockey, Richard Gregorski, and Ilse Lehiste

The theory of temporal compensation is based on the assumption that temporal programming information for "chunks" of speech larger than one linguistic segment is utilized at some unspecified, but rather late, level in the speech production mechanism. It is assumed that language is programmed in units no smaller than those defined by traditional manner-of-articulation parameters. Further, it is assumed that the domain over which temporal information is specified, and therefore over which the durational interaction described below takes place, is a programming unit.

This means that the duration of some multisegmental string of speech is fairly rigidly determined, and if this string or a stream of speech containing this string is repeated over many times at the same rate, the duration of the programming unit will remain very close to its average every time it is produced. But the same will not necessarily be true for the subparts of the programming unit. Since it is the duration of the higher-level unit which is predetermined, the durations of the individual segments are free to vary somewhat, as long as their sum approximates very closely the duration of the higher unit. The extent to which segments can vary is postulated to be determined by external factors such as whether or not segmental duration is contrastive in the language being considered.

Slis (1968) noted such a compensatory process in Dutch. He found that the lengths of several words of a given number of segments were quite similar despite substitution of segments with different intrinsic durations. A more sophisticated mathematical technique for testing for temporal compensation has been used by Kozhevnikov and Chistovich (1965) for Russian and by Ohala (1970), Allen (1969) and Lehiste (1970) for English.

The latter technique involves measurement of segments and determination of their variances and of correlation coefficients between adjacent segments and groups of segments. The assumption is that if there is little or no correlation in duration between adjacent segments, then at some level each segment is programmed separately. If so, the variance of the whole utterance or of any subpart of it should be equal to the sum of the variances of the individual segments. If an utterance is programmed in terms of more than one segment, we expect negative correlations between subparts of the largest programming unit; that is, if one part is longer than average, another part will be shorter than average to allow the duration of the programming unit to come quite close to its own average. If a negative correlation is found, it should also be true that the sum of the variances of the subparts of the utterance is greater than the variance of the programming "chunk."

In her 1970 experiment, Lehiste found that negative correlations exist between subparts of mono- and disyllabic words in English. The experiment to be reported was designed to discover whether temporal compensation operates within word-size units when they are included in a sentence.

Methods

Two subjects, PM and LS, both graduate students at The Ohio State University, were used. Each was seated in an I.A.C. sound-treated chamber with a high-quality Ampex microphone about one foot from his mouth. Cards with the utterances we wished to elicit written on them were placed on a table in front of the subject, one at a time. The subjects were asked to repeat a given utterance at a steady, comfortable rate of speech until signalled to stop. Each utterance was repeated 150 times or more. Recordings were made on an Ampex 350 magnetic tape recorder at a speed of 7 1/2 i.p.s.

One word, "Aggie," and one sentence, "I bag Aggie," were recorded by both subjects. In addition, speaker PM recorded the word "Agatha" and the sentence "I saw Agatha." These utterances were chosen on the basis of potential segmentability.

The recordings were then processed through a Frøkjaer-Jensen Trans-Pitch meter and recorded in the form of a duplex oscillogram by an Elema-Schönander Mingograf at a speed of 100 mm./sec. The oscillograms were segmented following the standards set forth in Naeser (1969). The duration of each segment was measured, with an accuracy estimated to be to the nearest 1/2 mm. or 5 msec.

Both Ohala (1968) and Kozhevnikov and Chistovich (1965) used normalization procedures involving choosing out of their total set of data a group of utterances of highly similar duration, to eliminate possible effects of differences in rate. Following their precedent, we have based our conclusions on the 50 utterances closest to the mean for each utterance and each speaker, in the belief that only when variability of duration of the entire utterance is carefully constrained can small variations within the utterance be examined meaningfully.

The results were processed by an IBM 360 computer. Statistical measures derived were mean duration, standard deviation, variance, relative variance ($\frac{V}{M}$), coefficient of variation ($\frac{\sigma}{M}$) and Pearson correlation coefficient $r = \frac{1}{N}\Sigma\left(\frac{x-\bar{x}}{\sigma x}\right)\left(\frac{y-\bar{y}}{\sigma y}\right)$.

Statistical tests were run on the following segments and combinations of segments: 1) individual segments with each other, 2) all possible combinations of n segments, with each other and with other combinations of n segments, where n ranges from 2 to the number of segments in the utterance minus one, and with the provision that the two sets being tested for correlation have no segments in common. When n > 2, only adjacent sets of three, four, etc. are used. 3) individual segments with sets of n segments. In addition, measurements were made of the pauses between the utterances and correlations were calculated between utterances and pauses.

Results

1. Of standardization.

We found that standardization of "rate" in this very restricted sense gave us a much clearer picture of which segments and combinations of segments were interacting with each other than we could have formed by looking at the complete set of 150 tokens. Following is a chart showing numbers of significant negative correlations found in the largest group and in the subset of 50 for the two sentences:

TABLE I

Numbers of significant negative correlations at the .01 level before and after normalization

		"I bag Aggie"	"I sass Agatha"
PM	50	55	94
	150	1	45
LS	50	42	
	150	12	

2. For words in isolation:

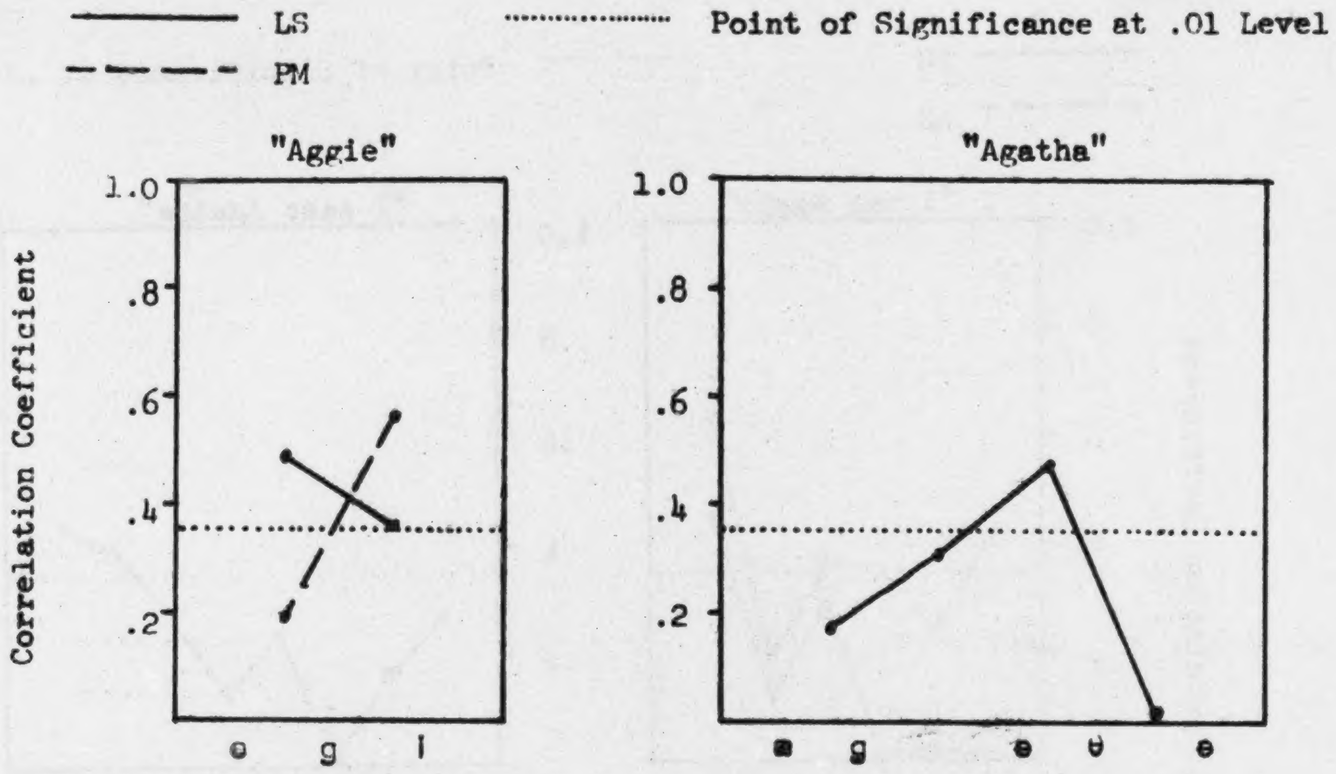
In the majority of cases, there were significant negative correlations (at the .01 level) between adjacent segments in the word "Aggie" for both speakers. Although negative correlations were present in all cases between adjacent segments in the word "Agatha" as spoken by PM, all except one were below the .01 level of significance. Higher negative correlations, predominantly significant, were found when larger portions of the word were tested, the highest negative correlation coefficient values being for mutually exclusive subsets of the whole word, e.g.[ægə-θə].

Typical results are presented graphically in Fig. 1. Tables containing additional information on mean, standard deviation, variance, relative variance, and correlation coefficient are to be found in the appendix.

3. For sentences:

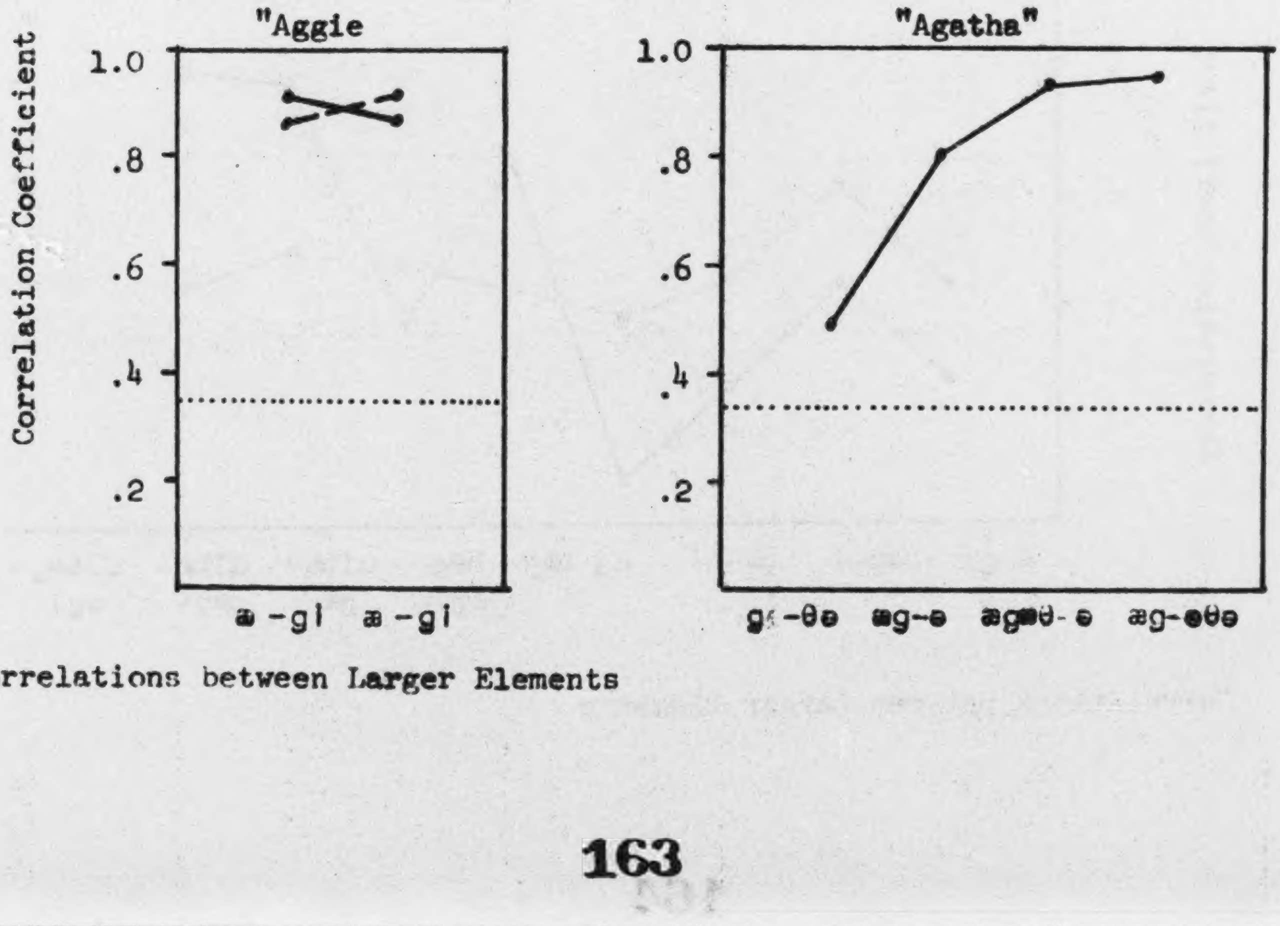
Correlations between adjacent segments in the sentences "I bag Aggie" and "I sass Agatha" were all similar to those for the word "Agatha", negative, but tending to be below the .01 level of significance. However, note in the following graph (Fig. 2) that for both speakers and both

Figure 1.



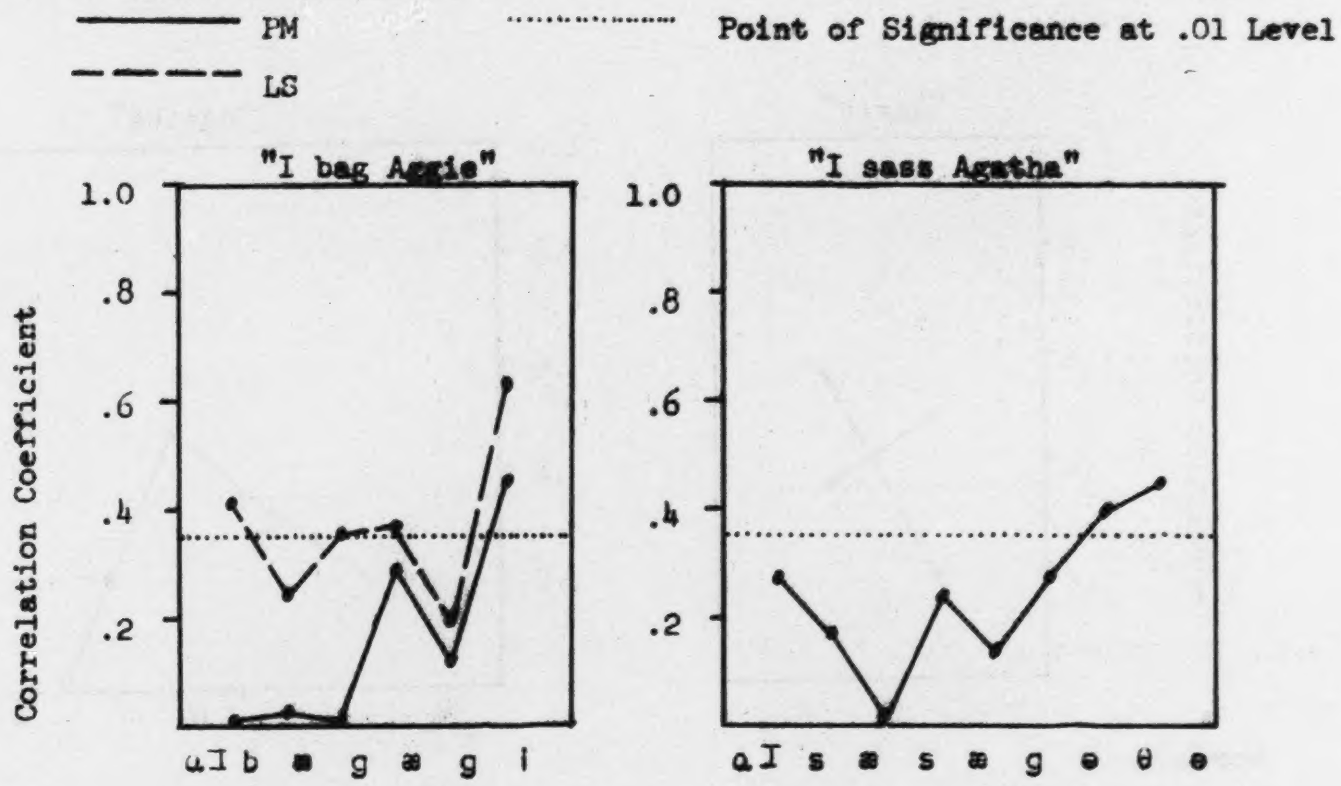
Words:

Correlations between Adjacent Segments



Correlations between Larger Elements

Figure 2.

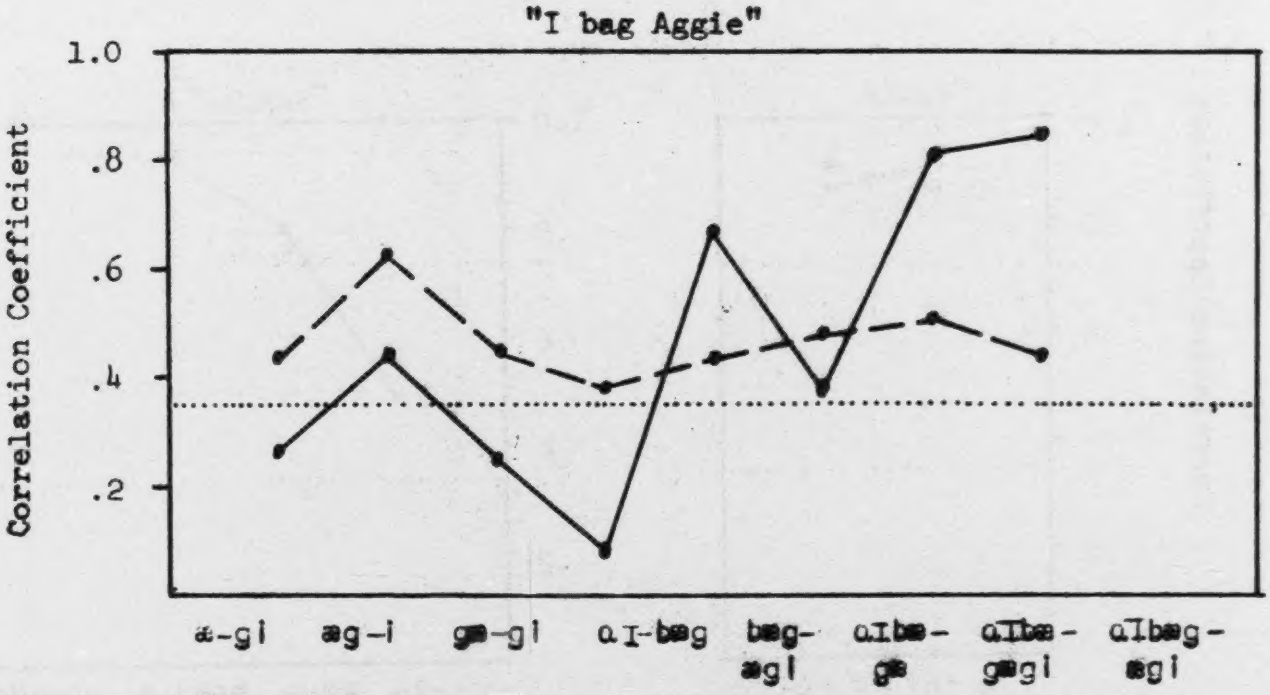


"I bag Aggie"

"I sass Agatha"

Sentences:

Correlations between Adjacent Segments



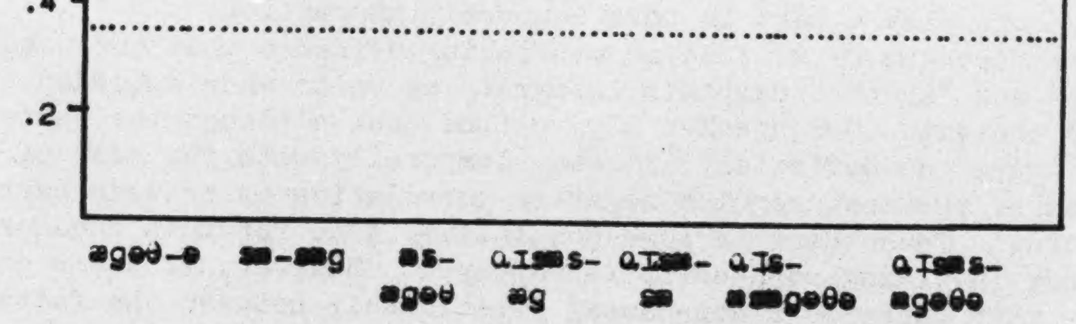
"I bag Aggie"

Correlations between Larger Elements

sentences there is quite a strong negative correlation between the last two segments (this may indicate a tendency for temporal adjustment to take place utterance-finally).

For speaker PM, there is a tendency to have stronger correlations between units of larger sizes, the largest being between mutually exclusive subparts of the whole sentence. This does not hold for LS, although her correlations between segments are consistently smaller on the average than her correlations between larger elements (see Figs. 2 and 3).

Figure 3.



For speaker PM there are two significant negative correlations between subparts of the word "Aggie" when it is included in the sentence "I bag Aggie," both in the low range (see Table 4, Appendix). For LS there were three (Table 5), two of them being the highest negative correlations for this speaker and this sentence. For speaker PM, the same tendencies hold for "Agatha" when it is included in the sentence "I sass Agatha"--there are seven significant negative correlations between elements of the word, but higher negative correlations result from testing larger portions of the sentence with no consideration for traditional word boundaries.

4. For utterance and pause:

As may be seen from the following table, very high negative correlation coefficients were found for all tested utterances compared with the following pause:

TABLE 2

Correlation coefficients for whole utterance and following pause.

Speaker	"Aggie"	"Agatha"	"I bag Aggie"	"I sass Agatha"
PM	-.878	-.820	-.943	-.901
LS	-.710		-.828	

Discussion:

The most obvious conclusion to be reached from these data is that, if we do indeed have a legitimate means of detecting temporal compensation in examining variation and correlation coefficient, temporal comepnsation occurs in a high degree between portions of these short utterances. It would appear, then, that at some level the entire utterance is programmed as a whole, since all segments and combinations of segments play a part in this temporal interaction.

For speaker PM we find no convincing evidence that the words "Aggie" and "Agatha" maintain integrity as units when embedded in a longer context. For speaker LS, we find that although the parts of the word "Aggie" do definitely interact temporally with the rest of the utterance, the most regular negative correlation is between parts of the word. Thus there is some possibility that for this speaker, a strategy involving word-units is employed. However, it seems equally likely that there is a non-causal relationship between the facts that there is a high negative correlation between [æg] and [i] for speaker LS and that [ægi] can be an utterance by itself. Further studies will be needed to disambiguate these data.

In the present study, lexical words did not emerge as units within which temporal compensation takes place. Rather, they seemed to be merged into a phonological phrase, losing their separate identity. It thus seems unlikely that the "word" level will prove as useful in phonological description as it does for lexical and syntactic description. It is not inconceivable that temporal compensation may serve to determine the extent of linguistic units at a higher level, such as a phrase or breath-group. The next step, of course, is to examine utterances containing an embedded sentence or phrase for units displaying internal cohesion.

Much research has been based on the hypothesis that the syllable is the basic unit of speech production, especially by Stetson (1951) and Kozhevnikov and Chistovich (1965). There is some evidence from electromyography that this may be so at some level in articulatory programming (MacNeilage and DeClerk, 1968). But our results show a singular lack of evidence for postulating either the CV or VC syllable as a basic unit of temporal programming for English. There are, for all of our sets of data spoken at a very similar rate, various degrees of negative correlation between most adjacent segments with no clear indication of a stronger bond between CV or VC sequences.

We agree with Ohala's statement that "Chistovich and her colleagues took the units [of speech production] to be syllables based on the results of a previous experiment, in which it was shown that the duration of the words and syllables relative to the duration of the whole utterance remained constant during changes of rate, but the relative durations of the consonants and vowels, the components of the syllable, varied during changes of rate. Thus the smallest interval maintaining relative temporal "integrity" in the face of changes in rate was the syllable--at least in Russian. But these results could as well be taken as indicating that the articulatory unit could be no smaller than the syllable but it could be larger (p. 145)."

While it is undeniable that the syllable plays a significant part in speech rhythm and may at some level be a measure of speech units, we find no evidence for postulating it as a primary building block, in English, at the level of programming which we presume to be observable through the process of temporal compensation.

The amazingly high negative correlations between the speech portions of our data and the following pauses reflect the high accuracy with which our subjects were able to execute the request to speak at a steady rate. We realize that the speech situation which we have created is artificial in that it is conducive to a measured rhythm; however we still feel that it is interesting to note that in all probability the pause is programmed with the speech as a temporal unit. The internal programming of the utterance itself apparently takes place at one level; at the next higher level, the unit of programming is the sentence plus the following pause. This may indicate, as suggested by Ohala (1970), that the mechanism for isochrony is indeed part of the linguistic competence of the speaker of English.

Acknowledgments

We would like to thank Dr. John Black of the Speech Department (O.S.U.) for the loan of equipment. Especially we wish to thank Thomas G. Whitney of the Ohio State University Instruction and Research Computer Center, without whose generous assistance this project would have been inconceivably complicated.

Bibliography

Allen, George. 1969. "Structure of Timing in Speech Production." Paper presented at the San Diego Meeting of the Acoustical Society of America, November 4.

Kozhevnikov, V.A., and L. A. Chistovich. 1965. Speech: Articulation and Perception. Translated by J.P.R.S., Washington, D.C. , No. JPRS 30, 543. Moscow-Leningrad.

Lehiste, Ilse. 1970. "Temporal Organization of Spoken Language," Working Papers in Linguistics No. 4, CIS Research Center, The Ohio State University, 95-114.

MacNeilage, Peter, and Joseph DeClerk. 1969. "On the Motor Control of Coarticulation in CVC Monosyllables." Journal of the Acoustical Society of America 45.1217-1233.

Naeser, Margaret. 1969. "Criteria for the Segmentation of Vowels on Duplex Oscillograms," Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin.

Ohala, John. 1970. "Aspects of the Control and Production of Speech." Working Papers in Phonetics 15, U.C.L.A.

Slis, I. H. 1968. "Experiments on Consonant Duration Related to the Time Structure of Isolated Words." I.P.O. Annual Progress Report, No. 3.71-80. Institute for Perception Research, Eindhoven, Holland.

Stetson, R. H. 1951. Motor Phonetics. Amsterdam.

Appendix

These tables are to be read as follows:

The (a) tables indicate the mean, standard deviation, variance, and relative variance for each of the variables to be used when testing for correlation.

The (b) tables show correlation coefficients, ordered from lowest to highest for each utterance. On either side of a hyphen are the two variables being considered. Notice that a variable may contain any number of segments and that the correlations represented are between the two variables on either side of the hyphen taken as units. Therefore, if you see ɑɪb-ægl, this means we are considering [ɑɪb] as a unit in this particular comparison and plotting its durational values against those of the "unit" [ægl].

Since the means, standard deviations, etc. of the sum of the items being compared are always identical with the same information for one of the variables when we are dealing with words [æ+gl = ægl] and since the same is often true when we are dealing with sentences [ɑɪb- æg= ɑɪbæg] this information is left ouf of the table when there is overlap.

Notice that the comparisons whose sum is not equal to one of the variables involve non-adjacent elements. About 1/2 of PM's significant negative correlations are for non-adjacent units, but for LS only two are. This may be further evidence for a difference in programming strategy.

Formulae for statistical variables are:

$$r = \frac{1}{N} \sum \left(\frac{x-\bar{x}}{\sigma x}\right)\left(\frac{y-\bar{y}}{\sigma y}\right)$$

Mean (M): $\frac{x}{N}$

Standard deviation: $\sigma = \sqrt{\frac{\sum (x^2 - \bar{x}^2)}{N}}$

Variance: $V = \sigma^2$

Relative Variance: $= \frac{V}{M}$

Variation coefficient: $= \frac{\sigma}{M}$

TABLE 3

Speaker PM: "Aggie"

	r	M	σ	σ^2	Relative Variance	Variation Coefficient
(a) Variable						
æ		169.20	13.05	170.36	1.007	.077
g		64.80	11.44	130.96	2.021	.177
i		169.20	11.63	135.36	.800	.069
æg		234.00	11.09	123.00	.526	.047
gi		234.00	13.08	171.00	.731	.056
æ+i		338.40	13.29	176.56	.522	.039
ægi		403.20	5.81	33.81	.084	.014
(b) Variables						
g-i	-.358					
æ-i	-.425					
æ-g	-.597					
æg-i	-.870					
æ+i-g	-.900					
æ-gi	-.901					

TABLE 4

Speaker LS: "Aggie"

	r	M	σ	σ^2	Relative Variance	Variation Coefficient
(a) Variable						
æ		237.40	16.47	271.25	1.143	.069
g		89.30	14.42	208.02	2.329	.162
i		165.70	19.93	397.01	2.396	.120
æg		326.70	19.74	389.69	1.193	.060
gi		255.00	16.82	283.00	1.110	.066
æ+i		403.10	16.74	280.06	.695	.042
ægi		492.40	8.74	76.38	.155	.018
(b) Variables						
g-i	-.560					
æ-i	-.591					
æ+i-g	-.853					
æ-gi	- 862					
aæg-i	-.903					

TABLE 5

Speaker PM: "Agatha"

	r	M	σ	$σ^2$	Relative Variance	Variation Coefficient
(a) Variables						
æ		100.50	15.21	231.25	2.301	.151
g		45.40	9.64	92.84	2.045	.212
ə		67.50	9.45	89.25	1.322	.140
θ		92.70	10.78	116.21	1.254	.116
ə		140.50	16.13	260.25	1.852	.115
æg		145.90	16.58	271.70	1.862	.113
gə		112.90	11.18	125.09	1.108	.099
əθ		160.20	10.39	107.96	.674	.065
θə		233.20	18.89	356.76	1.530	.081
ægə		213.40	17.65	311.45	1.459	.083
gəθ		205.60	12.64	159.64	.776	.061
əθə		300.70	17.06	291.06	.968	.057
ægəθ		306.10	16.75	280.44	.916	.055
ægəθə		446.60	5.71	32.63	.073	.013
(b) Variables						
ægə–θ	-.387					
ə$_1$–θə	-.434					
gə–θə	-.470					
ə$_1$–θ	-.479					
æ–ə$_2$	-.757					
æ–θə	-.761					
æg–θə	-.772					
æg–ə$_2$	-.810					
æ–əθə	-.828					
ægə–ə$_2$	-.858					
ægəθ–ə	-.940					
æg–əθə	-.943					
ægə–θə	-.953					

TABLE 6

Speaker PM: "I bag Aggie"

	r	M	σ	σ^2	Relative Variance	Variation Coefficient
(a) Variables						
ɑɪ		100.70	7.75	60.01	.596	.077
b		70.40	6.47	41.84	.594	.092
æ		136.90	10.58	111.89	.817	.077
g		51.10	6.02	36.29	.710	.118
æ		159.40	11.94	142.64	.895	.075
g		56.30	10.09	101.81	1.808	.179
i		150.30	12.51	156.41	1.041	.083
ɑɪb		171.10	10.06	101.30	.592	.059
bæ		207.30	11.71	137.22	.662	.057
æg		188.00	12.12	147.00	.782	.064
gæ		210.50	11.67	136.25	.647	.055
æg		215.70	14.66	215.01	.997	.068
gi		206.60	11.98	143.45	.694	.058
ɑɪbæ		308.00	14.14	200.00	.649	.046
ɑɪbæg		359.10	15.74	247.81	.690	.044
ɑɪbægæ		518.50	14.53	211.25	.407	.028
ɑɪbægæg		574.80	15.59	243.19	.423	.027
ɑɪbægægi		725.10	8.61	74.06	.102	.012
bæg		258.40	13.13	172.50	.668	.051
bægæ		417.80	14.37	206.38	.494	.034
bægæg		474.10	15.36	235.88	.498	.032
bægægi		624.40	11.23	126.06	.202	.018
ægæ		347.40	14.98	224.31	.646	.043
ægæg		403.70	16.40	268.88	.666	.041
ægægi		554.00	12.17	148.00	.267	.022
gæg		266.80	13.89	192.88	.723	.052
gægi		417.10	13.50	182.25	.437	.032
ægi		366.00	14.49	210.00	.574	.040
(b) Variables						
g–ægi	-.367	417.10	13.50	182.25	.437	.032
ɑɪ–æ$_2$	-.378	260.10	11.52	132.63	.510	.044
ɑɪbæ-gæ	-.379					
æg$_1$-i	-.381	338.30	13.70	187.75	.555	.041
ɑɪbæ-gæg	-.381					
ɑɪbæ–æg	-.389	523.70	15.93	253.88	.485	.030
ɑɪbæ–æ	-.403	467.40	14.37	206.38	.442	.031
b–ægægi	-.405					
ɑɪb-gæg	-.408	437.90	13.42	180.19	.411	.031
ɑɪbæg-i	-.415	509.40	15.52	240.75	.473	.030
æ$_1$-i	-.416	287.20	12.58	158.25	.551	.383
bægæ-i	-.419	568.10	14.56	212.06	.373	.026
ɑɪb–æg	-.428	386.80	13.78	189.94	.491	.036

TABLE 6 (continued)

	r	M	σ	σ^2	Relative Variance	Variation Coefficient
(b) Variables						
ɑɪb_gæ	-.430	381.36	11.69	135.56	.358	.031
ægæ_l	-.434	497.70	14.78	218.31	.439	.030
æg_2-l	-.440	366.00	14.49	210.00	.574	.040
ɑɪbæ_l	-.446	458.30	14.10	198.81	.434	.031
bæ_gl	-.447	413.90	12.46	155.38	.375	.030
g_2l	-.445	206.60	11.98	143.45	.694	.058
bæg-gl	-.467	465.00	13.00	169.00	.363	.028
æ$_1$-ægl	-.472	502.90	13.31	177.25	.352	.026
ɑɪb_æ$_2$	-.473	330.50	11.41	130.25	.394	.035
ɑɪb^{æ}g-æg	-.476					
ɑɪbæg_æ	-.477					
æ$_1$-gl	-.482	343.50	11.54	133.25	.388	.034
æg_1-gl	-.494	394.60	12.12	147.00	.373	.031
æ-gægl	-.513					
ɑɪbæ-gl	-.539	514.60	12.69	161.00	.313	.025
ɑɪbæg-gl	-.541	565.70	13.68	187.13	.331	.024
ɑɪ-gægl	-.549	517.80	11.29	127.38	.246	.022
bæ-ægl	-.564	573.30	12.48	155.88	.272	.022
ɑɪ-ægl	-.567	466.70	11.95	142.69	.306	.026
ɑɪbægæ-l	-.579	668.80	12.53	157.06	.235	.019
æg-ægl	-.595					
ɑɪb-gægl	-.596	588.20	11.03	121.63	.207	.019
ɑɪ-ægægl	-.596					
ɑɪb-ægl	-.597	537.10	11.72	137.25	.256	.022
ægæ-gl	-.613					
ɑɪ-bægægl	-.644					
bægæ-gl	-.650					
bæg-ægl	-.674					
ægæg-l	-.676					
bægæg-l	-.693					
ɑɪb-ægægl	-.716					
ɑɪbæ-ægl	-.783					
ɑɪbægæ-gl	-.806					
ɑɪbæ-gægl	-.807					
ɑɪbægæg-l	-.835					
ɑɪbæg-ægl	-.841					

TABLE 7

Speaker LS: "I bag Aggie"

	r	M	σ	σ^2	Relative Variance	Variation Coefficient
(a) Variables						
aɪ		145.10	12.67	160.50	1.106	.087
b		58.60	7.55	57.04	.973	.129
æ		169.10	15.80	249.70	1.477	.093
g		57.90	9.17	84.09	1.452	.158
æ		197.20	12.50	156.16	.792	.063
g		80.30	12.06	145.41	1.811	.150
i		127.70	13.94	194.21	1.521	.109
aɪb		203.70	11.74	137.81	.677	.058
bæ		227.70	15.75	248.21	1.090	.069
æg$_1$		227.00	15.10	228.00	1.004	.067
gæ		255.10	12.47	155.50	.610	.049
æg$_2$		277.50	15.60	243.25	.877	.056
gi		208.00	11.36	129.00	.620	.055
aɪbæ		372.80	16.35	267.31	.717	.044
aɪbæg		430.70	14.94	223.06	.518	.035
aɪbægæ		627.90	15.11	228.44	.364	.024
aɪbægæg		708.20	17.19	295.38	.417	.024
aɪbægægi		835.90	14.73	216.94	.260	.018
bæg		285.60	14.17	200.75	.703	.050
bægæ		482.80	16.53	273.38	.566	.034
bægæg		563.10	16.82	288.06	.503	.030
bægægi		690.80	14.37	206.56	.299	.021
ægæ		424.20	17.16	294.44	.694	.040
ægæg		504.50	17.50	306.25	.607	.035
ægægi		632.20	16.06	257.94	.408	.025
gæg		335.40	14.86	220.94	.659	.044
gægi		463.10	12.57	158.06	.341	.027
ægi		405.20	12.81	164.00	.405	.032
(b) Variables						
aɪ-bæ	-.354					
aɪb-ægæg	-.362					
bæg-æg$_2$	-.364					
æ$_1$-g$_1$	-.365					
aɪbæg-æg$_2$	-.367					
b-æg$_1$	-.370					
g$_1$-æ$_2$	-.370					
g$_1$-æg$_2$	-.372					
æ$_1$-gægi	-.377					
g$_1$-ægi	-.383					
aɪ-bæg	-.385					
aɪbæ-gæg	-.397					

TABLE 7 (continued)

	r	M	σ	σ^2	Relative Variance	Variation Coefficient
(b) Variables						
bæ-gæg	-.397					
ɑɪb-æg	-.403					
ɑɪbæg-æ	-.404					
ɑɪbægæ-gl	-.409					
ɑɪ-bægæg l	-.412					
ɑɪ-b	-.417					
ægæ-gl	-.425					
æ$_2$-gl	-.427					
ɑɪbæ-g	-.428					
bæg-ægl	-.436					
ɑɪbæg-ægl	-.445					
gæ-gl	-.446					
b-ægægl	-.447					
bæ-gl	-.455	285.60	14.17	200.75	.703	.050
ɑɪb-ægægl	-.474					
ɑɪbæ-gæ	-.477					
ɑɪ-bægæ	-.490					
ægæg-l	-.497					
bæ-gægl	-.504					
ɑɪb-ægæ	-.506					
ɑɪbæ-gægl	-.507					
bægæ-gl	-.521					
ɑɪbægæ-l	-.569	835.90	14.73	216.94	.260	.018
bægæg-l	-.577					
gæg-l	-.621					
g$_2$-l	-.627					
æg$_2$-l	-.629					

TABLE 8

Speaker PM: "I sass Agatha"

	r	M	σ	σ^2	Relative Variance	Variation Coefficient
(a) Variables						
ɑɪ		103.90	8.56	73.29	.705	.082
s		94.20	5.95	35.36	.375	.063
æ		125.00	10.00	100.00	.800	.080
s		82.10	5.39	39.09	.354	.066
æ		134.00	8.49	72.00	.537	.063
g		43.00	7.21	52.00	1.209	.168
ə		57.40	8.14	66.24	1.154	.142
θ		82.10	9.60	92.09	1.122	.117
ə		137.70	11.23	126.21	.917	.082
ɑɪs		198.10	8.99	80.89	.408	.045
æs		207.10	10.91	119.10	.575	.053
æg		177.00	10.34	107.00	.605	.058
gə		100.40	9.27	85.84	.855	.092
əθ		139.50	9.81	96.25	.690	.070
θə		219.80	11.04	121.97	.555	.050
sæ$_1$		219.20	10.65	113.36	.517	.049
sæ$_2$		216.10	8.90	79.30	.367	.041
sæs		301.30	12.20	148.94	.494	.041
ægəθə		454.20	12.59	158.44	.349	.028
ɑɪsæs		405.20	13.27	176.00	.434	.033
ɑɪægəθə		558.10	12.09	146.06	.262	.022
sæsægəθə		755.50	9.06	82.13	.109	.012
ɑɪsæsægəθə		859.40	5.04	25.38	.030	.006
ɑɪsæ		323.10	12.08	146.00	.452	.037
ɑɪsæs		405.20	13.27	176.00	.434	.033
ɑɪsæsæ		539.20	10.32	106.44	.197	.019
ɑɪsæsæg		582.20	11.70	137.00	.235	.020
ɑɪsæsægə		639.60	11.36	129.00	.202	.018
ɑɪsæsægəθə		721.70	12.00	144.06	.200	.017
sæsæ		435.30	10.42	108.63	.250	.024
sæsæg		478.30	11.74	137.81	.288	.025
sæsægə		535.70	11.83	140.06	.261	.022
sæsægəθ		617.80	14.14	199.88	.324	.023
sæsægəθə		755.50	9.06	82.13	.109	.012
æsæ		341.10	9.61	92.44	.271	.028
æsæg		384.10	10.04	100.81	.262	.026
æsægə		441.50	11.32	128.25	.290	.026
æsægəθ		523.60	12.82	164.25	.314	.024
æsægəθə		661.30	8.16	66.63	.101	.012
sæg		259.10	11.65	135.81	.524	.045
sægə		316.50	12.18	148.25	.468	.038

TABLE 8 (continued)

	r	M	σ	σ^2	Relative Variance	Variation Coefficient
(a) Variables						
sægəθ		398.60	13.57	184.19	.462	.034
sægəθə		536.30	11.88	141.06	.263	.022
ægə		234.40	12.31	151.64	.647	.053
ægəθ		316.50	13.65	186.25	.588	.043
gəθə		320.20	9.95	99.00	.309	.031
əθə		277.20	10.16	103.19	.372	.037
(b) Variables						
sæ$_1$-ægəθ	-.356	535.70	14.00	196.06	.366	.026
ɑɪsæ-θə	-.362	542.90	13.09	171.25	.315	.024
sæsæg-θə	-.363	698.10	12.87	165.69	.237	.018
s$_2$-ə$_1$	-.365	139.50	7.95	63.25	.453	.057
ɑɪ-sæsæg	-.369					
ɑɪ-sægəθ	-.375	502.50	13.05	170.25	.339	.026
ɑɪsæs-θ	-.380	625.00	13.66	186.50	.298	.022
æs-əθə	-.384	484.30	11.72	137.25	.283	.024
g-əθə	-.384					
ɑɪs-æsæ	-.387					
ɑɪsæsæg-əθ	-.388					
ɑɪsæsæg-ə$_1$	-.390					
sæsæ-ə$_2$	-.392	573.00	11.96	143.00	.250	.021
ɑɪs-æsægə	-.393					
ə$_1$-θ	-.397					
ɑɪ-ægəθə	-.397	558.10	12.09	146.06	.262	.022
æ$_1$-ægəθ	-.403	441.50	13.28	176.25	.399	.030
sæ$_1$-gəθə	-.404	539.40	11.26	126.75	.235	.021
s$_2$-ɑɪ +ægəθə	-.408	640.20	11.04	121.94	.190	.017
ɑɪs-sægəθə	-.412	734.40	11.57	133.88	.182	.016
sæg-əθə	-.414					
ɑɪ-sæsægə	-.416					
ɑɪsæsæ-ə$_2$	-.419	676.90	11.64	135.50	.200	.017
s$_2$-əθə	-.421	359.30	9.28	86.19	.240	.026
sæs-əθə	-.423	578.50	12.13	147.25	.255	.021
ɑɪ-sæsæ	-.423					
ɑɪs-əθə	-.424	475.30	10.33	106.63	.224	.022
æ$_1$-gəθə	-.437	445.20	10.59	112.06	.252	.024
ɑɪs-æsægəθ	-.438					
æ$_1$-sægəθ	-.442					
sæ-sæ	-.443	359.40	11.91	141.75	.394	.033
æ$_1$-ægə	-.446	603.90	11.13	123.81	.205	.018
æsæg-θə	-.446					
θ-ə$_2$	-.447					
sæ-sæg	-.449					
ɑɪ-sægəθə	-.454	640.20	11.04	121.94	.190	.017
æ$_1$-æ$_2$	-.454	259.00	9.75	95.00	.367	.038

TABLE 8 (continued)

	r	M	σ	σ^2	Relative Variance	Variation Coefficient
(b) Variables						
sæs-gəθə	-.458	621.50	11.69	136.75	.220	.019
ɑɪ-æsægəθ	-.460	627.50	11.68	136.50	.218	.019
æsæg-ə$_2$	-.461	521.80	11.09	123.00	.236	.021
sæsæg-ə$_2$	-.469	616.00	11.85	140.44	.228	.019
sæ$_1$-sægə	-.469					
ə$_1$-θə	-.473					
æs-ægəʊ	-.474					
sægə-ʊə	-.480					
sæ$_1$-æg$_2$	-.487	453.60	11.71	137.19	.302	.026
ɑɪsæsæg-ə$_2$	-.488	719.90	11.61	134.88	.187	.016
æ$_1$-sægə	-.493					
æsægə-ə$_2$	-.499	579.20	11.29	127.44	.220	.019
sæ-æ$_2$	-.501	353.20	9.74	94.81	.268	.016
ægəθ-ə$_2$	-.502					
æs-gəθə	-.514	527.30	10.32	106.44	.202	.020
æs-ægə	-.530					
ɑɪsæs-æg	-.532					
æs-æ$_2$	-.533					
ɑɪsæsægəθ	-.533					
əθ-ə$_2$	-.541					
sæs-æ$_2$	-.542					
sæsægə-ə$_2$	-.552	673.40	10.93	119.50	.177	.016
ɑɪsæ-sæ$_2$	-.553					
sægəθ-ə$_2$	-.555					
ɑɪsæ-sægə	-.562					
ɑɪsæ-sægəθ	-.568					
æ$_1$-sæg	-.579					
ɑɪsæ-æg	-.581	500.10	10.38	107.69	.215	.021
ɑɪsæ-æ$_2$	-.584	457.10	9.91	98.25	.215	.022
ɑɪsæsægə-ə$_2$	-.587	777.30	10.27	105.50	.136	.013
ɑɪsæs-ægəʊ	-.603					
æ$_1$-æg	-.604	302.00	9.06	82.00	.272	.030
sæsæ-gəθə	-.605					
ɑɪsæs-æg	-.608					
ɑɪsæs-əθə	-.609	682.40	10.72	115.00	.169	.016
ɑɪsæs-æ$_2$	-.629					
ɑɪsæs-gəʊə	-.639	725.40	10.31	106.31	.147	.014
ɑɪsæ-gəθə	-.652	661.30	8.16	66.63	.101	.012
æ$_1$-ægəθə	-.653	579.20	9.70	94.13	.163	.017
æsæg-əʊə	-.673					
sæ-sægəθə	-.681					
sæsægə-ʊə	-.688					
sæs-ægəʊə	-.733					
æ$_1$-sægəθə	-.734					
æsægə-ʊə	-.734					

TABLE 8 (continued)

	r	M	σ	σ^2	Relative Variance	Variation Coefficient
(b) Variables						
$æ_1$-αɪ +ægeϑe	-.736	683.10	8.25	68.06	.100	.012
αɪ-æsægeϑe	-.759	765.20	5.81	33.81	.044	.008
sæsægeϑ-e	-.768			41.81		
æsægeθ-e	-.777					
αɪs-æsægeϑe	-.832					
αɪ-sæsægeϑe	-.838					
αɪsæ-ægeθe	-.863	777.30	6.47	41.81	.054	.008
αɪsæsæ-geθe	-.877					
αɪsæsæg-ϑe	-.899	859.40	5.04	25.38	.030	.006
αɪsæsæg-eθe	-.903					
αɪsæsægeθ-e	-.908					
αɪsæ-sægeθe	-.912					
αɪsæs-ægeθe	-.925					

Relative Intelligibility of Five Dialects of English

Mary Virginia Wendell

Relative Intelligibility of Five Dialects of English

Mary Virginia Wendell

In the production of the various linguistic Atlases of the English language, numerous word lists and phonetic descriptions have been made of the many regional and social dialects of English, with "dialect" being defined as special varieties of usage and/or pronunciations within the range of a given linguistic system. (Reed, 1967, p. 2) Thus, a language may be considered to be a collection of related dialects in a particular area, often encompassing a single nationality. Carroll Reed (1967, p. 2) has said, "Languages are not mutually intelligible; different dialects of the same language are ordinarily mutually intelligible (with some notable exceptions, such as certain dialects of Chinese)." The purpose of this study is to determine how intelligible certain dialects of English are to native speakers of one particular dialect.

In a study by L. S. Harms (1961), listeners of three status groups attempted to reconstruct spoken messages of speakers of the three statuses. Listeners achieved highest comprehension scores when speaker and listener status were the same. In the present study, this result has been modified to include unintelligibility of regional dialects. Five dialects of English were presented to listeners who were native speakers of one of the dialects. Highest intelligibility scores were expected when speaker and listener dialect coincided. Since no other data in relative intelligibility of the five dialects involved in the study are available, no prediction was made as to the most difficult dialect to understand.

Dialects for the study were chosen on the basis of their differences from the control dialect, which was that of Columbus, Ohio. At least two of the speakers chosen demonstrated idiolectal differences, but the speakers were selected because their speech patterns were very close to those of the dialects they represented, and quite different from those of the control dialect.

The dialects chosen for the study were: Columbus, Ohio, an example of General American speech; Long Island, New York, Jewish community; Portsmouth, Ohio, an example of what can be called Rural Southern Ohio speech, a mixture of General American and Southern speech; one variety of British stage speech; and Black American, (urban variety of this dialect, rather than what is known as Southern Negro speech). No attempt was made to investigate intelligibility of dialectal words and sentence patterns. The test which was used examined only word intelligibility, i.e. pronunciation differences.

Brief descriptions of the dialects follow. All are taken from C. M. Wise's Applied Phonetics (1958). Only the more prominent features are listed with particular attention to those characteristics which are applicable to the listening test items.

General American Speech is characterized by the following pronunciations:

1. [ɔ] is the most common low back vowel except in words like *water*, *sorrow*, *not* and *possible*, where the vowel is [ɑ].
2. [æ] and [ɛ] are in free variation in words like *care* and *dear*.
3. Stressed long vowels diphthongise; for example, [eɪ] in *ate* and *stay*; [ou] in *go*, *soul*, and *below*. The diphthongs appear as pure vowels in weakly stressed syllables.
4. Central vowels are [ʌ], which is close to [ə] except in tenseness and duration, and [ɝ], as in *bird*, *turn*, and *murmur*.
5. All unstressed vowels reduce to [ə] or [ɪ] except [e] and [i] before another vowel. These two vowels reduce to [i]. [ə] before [l], [m], [r], and [n] reduces to syllabic [l̩], [m̩], [r̩], and [n̩].
6. [r] is always pronounced and never intrusive except sometimes in *wash*.
7. [l] is usually back except after high front vowels, and is often rounded after rounded vowels.
8. [t] is frequently lost when final.

The Southern-General American border region is characterized more by stress and intonation patterns than by specific phonetic qualities, but some characteristics are evident:

1. Retracted stress is common in words like *cement* and *insurance*.
2. Words are frequently run together and forms like *you'ns*, *you'zl*, and *y'all* are common for *you* (plural), *you will* and *you all*, respectively.
3. [ɛ] goes to [ɪ] always before nasals except in *been* and *since*, where the opposite happens.
4. [ɪ], [ɛ], and [æ] are raised before all front consonants.

Black Urban speech is characterized by voice quality as much as any other factor, but a few outstanding phonetic tendencies are indicated:

1. Word final stops are nearly always lost.
2. [θ] goes to [t] and [ð] to [d], particularly in pronouns and demonstratives.

3. Stressed vowels diphthongize, and the resulting diphthong sounds very like the first element.

4. Voiced consonants are often substituted for unvoiced ones; the reverse situation occurs equally often.

5. Consonant clusters are simplified usually by deletion of the stop in syllable final [sk], [sp], and [st] clusters.

The speech of New York City varies from borough to borough within the city. Some characteristics of the speaker from Long Island are listed here:

1. [ɔə] appears whenever a low back rounded vowel is followed by [r] as in horse.

2. Unrounded back vowels followed by [r] are lengthened as in New England speech, and the [r] is deleted.

3. [æ] is in free variation with [ɛə].

4. In nasalization, the nasal consonants are absorbed by the preceding vowels.

5. [ŋg] occasionally alternates with [ŋ] as in Long Island.

6. [|] is back and palatalized, often with no contact between tongue and alveolar ridge.

The variety of British speech used in the study has been somewhat Americanized, but still retains the "clipped" quality of British speech, and has a variety of low back vowels, most of which are not heard in General American speech.

1. Unstressed vowels reduce to [ɪ].

2. [æ] usually occurs in words like carry and parry; [ɛ] occurs in monosyllabics with [r].

3. [ɑ] is the so-called "broad a" in bath, half, aunt, etc.

4. [ɔ] is somewhat higher than American [ɔ], suggesting [o] when followed by [r], [|], and [w].

5. [ɜ] occurs in words like bird, turn and murmur; final [r] goes to [ə].

6. [ɾ] occurs intervocalically.

7. [l] is clear and frontal.

The selection of the testing procedure presented the greatest problem. A test was desired which perceived the different dialectal intonations, yet tested the intelligibility of specific words. The large number of listeners necessitated a test which could be easily scored. Tests in which the listeners write down their answers, whether sentences, words, or nonsense syllables, involve a degree of phonetic sophistication and judgment on the part of both participant and scorer, particularly if the experimenter is interested in what errors occur.

Phonetically balanced word lists such as the Harvard PB Lists and various CVC word lists were unsuitable because intonation patterns are lost when the speaker pronounces one word at a time. Fairbanks' Rhyme Test and the Modified Rhyme Test, developed by House, et al., are multiple choice tests where the alternative responses differ from the pronounced word by one phoneme. These tests, while eliminating the need for judgment in scoring, still present the problem of single word utterances which are inadequate for testing dialect intelligibility. A problem also arises because listeners only have a choice between four or five expected responses.

The Cloye Procedure test used in Harms' study presents a form to the listener on which a short narrative, heard previously, is printed with blanks replacing certain words. The subject is instructed to fill in the blanks with the exact word used by the speaker. This kind of test has listener comprehension as its main parameter, rather than auditory intelligibility.

The test selected for the study was the Multiple Choice Intelligibility Test, developed by Haugen, Black, et al. (1963). These tests are constructed of twelve lists of twenty-four words each. There are four forms, A, B, C, and D, and four alternate response forms, A-1, B-1, C-1, and D-1. Words are separated into groups of three words with a carrier phrase, pronounced with no pause, as if it were an incomplete sentence. The carrier phrase is the number of the test item, with eight items per each of the twelve lists in one test. Thus, the first item would look like:

Number 1 crook fair amble

The answer sheet includes four possible responses for each of the three words and the listener is asked to consider each word and make the correct response.

The methodology of the test, i.e. the fact that each item of seven or eight syllables is read as a phrase, preserves the intonation and assimilation tendencies of each dialect, yet provides an exact measure of word intelligibility. Each word in a particular utterance is scored separately; analyses of variance have shown little or not difference among the three scores (Black, 1958). Because a multiple choice format specifies possible responses, the importance of linguistic sophistication among the listeners is reduced, and the study of confusion characteristics between the fixed population of words is made possible. The limitations which result from fixed responses are counterbalanced by the need for a test in which phonetic knowledge is not necessary.

The twelve lists of each test contain different words, but are equivalent in difficulty. Equivalent but unlike lists are necessary to prevent a learning factor from affecting the reliability of the test as a measure of intelligibility. Forms A and B are somewhat less difficult in that they yield higher mean scores than Forms C and D. Form A was chosen for this study because of the naivety of high school age listeners which were employed.

Methodology

Six speakers recorded two lists of twenty-four words using an Ampex Model 350 tape recorder at 7 1/2 i.p.s. One list was taken from Form A of Black's Multiple Choice Intelligibility Tests; the second list was taken from the alternate response Form A-1. The lists of possible responses are identical for both forms; correct responses are different for each form. The lists were recorded in order that no speaker would read a list and its alternate (Speaker 6 was added after the recordings were finished, so that, in fact, he recorded two similar lists).

SPEAKER	DIALECT	LIST NO. A	LIST NO. A-1
1. C.B.	Columbus, Ohio	1	2
2. M.G.	New York-Jewish	2	3
3. B.N.	Rural Ohio	3	4
4. G.D.	British	4	5
5. J.H.	Columbus, Ohio	5	1
6. C.D.	Urban Black	6	6

The recordings were played on a Tandberg Model tape recorder to 65 senior high school students from four church groups located in the north side of Columbus, Ohio. The recordings were played in small meeting rooms with normal "classroom quite," with no noise in the signal. Listeners recorded their responses on standardized, printed answer sheets (Appendix 2), which had been duplicated by Multilith from the booklet "Multiple Choice Intelligibility Tests." Instructions for the listeners were adapted from the same booklet. The answer sheets were scored and checked by another scorer, and a frequency count of all litenener responses was done. Per cent counts were used to show how frequently each possible response was marked. Percentages were calculated by means of a simple Fortran program for an IBM 360 computer. The table was based on 63 as 100%, which was the number of usable listener responses for each list. Mean scores and standard deviation were calculated by computer.

Data analysis was performed on the basis of variance of mean intelligibility scores between dialects, using the Columbus speakers as controls, and assuming the mean scores of the control dialect to be 100% intelligible. Actual deviations from 100% intelligibility were assumed to be functions of the testing procedure.

Results

The results of the experiment are shown in Lists 1 through 6. The possible responses are shown on the left (N.A. indicates no answer was given). The numbers are the percentages of listeners who indicated each response. Correct responses are underlined. Since answer sheets for both Forms A and A-1 are the same, two compilations are shwon on each list. The speaker who read each list is shown by initials at the top. A percentage conversion chart is shown in Appendix 1 indicating the percentage of 63 versus the number of listeners.

Mean scores for each dialect are shown below--the average number correct out of 48. Scores are shown in order of most intelligible to least intelligible to listeners from Columbus, Ohio.

Speaker J.H. - Columbus - 45.24
Speaker C.B. - Columbus - 43.72
Speaker G.D. - British - 42.13
Speaker C.D. - Black - 39.83
Speaker M.G. - New York - 39.03
Speaker B.N. - Rural Ohio - 35.86

Pages 1 and 2 of each listener's test form were separated for ease in scoring so mean scores and standard deviations were calculated for each speaker's lists separately. In the table below two scores are shown for each speaker; the upper score is from the list on Form A, the lower from Form B.

SPEAKER	LIST NUMBER	MEAN	S.D.
J.H.	5	21.79	2.06
	1	23.44	0.86
C.B.	1	20.78	3.40
	2	22.60	2.20
G.D.	4	25.65	3.24
	5	21.46	6.95
C.D.	6	20.38	1.93
	6	19.44	1.82
M.G.	2	19.05	2.10
	3	19.97	1.82
B.N.	3	15.24	3.49
	4	18.97	3.68

It was noted that scores for the alternate response form A-1 were slightly higher than those of form A. This was not predicted in the preparation of the test materials, and both forms were combined in the calculation of the overall mean scores.

Since no test of significance for percentages in groups of four could be located, any deviation over 15% (10 listeners of 63) will be considered in the analysis. Since some of the words on the test are easily confused in standard testing situations, some of these differences will not be explainable in terms of dialect differences, but rather as perceptual confusions inherent in the words and their alternate responses.

The first Columbus speaker, J.H., shows only a few instances where less than 85% of the listeners responded correctly. In all but three cases, the confusions are between stops, or between stops and ø, as between word, were; plot, clock, blot; kind, pine, time; quit, quick; world, whirl.

Trial was mistaken for *trail* 15.87% of the time. The only plausible explanation for confusion between [aɪ] and [eɪ] would be that the listeners were mistaken in orthography. The speaker pronounced *trial* very clearly and the experimenter can find no phonetic basis for the confusion.

Relieve was mistaken for *relief* 19.05% of the time. [v] and [f] in final position are commonly confused, and since *relief* is the final word of the utterance, the drop in volume would augment this tendency.

Legion was mistaken for *legend* by 22.22% of the listeners. In the test item, *legend* is followed by *blunder*, nearly obscuring the [d], if, indeed, it was pronounced at all. *Legion-legend* shows a tense-lax apposition which is confused in many Ohio pronunciations as in /mɛz / and /mež /.

The errors indicated for the other Columbus speaker, C.B., are somewhat more complicated. *Court* was mistaken for *quart* nearly half the time. C.B.'s [w] in *quart* was unvoiced, and nearly imperceptible. Instead of a clear [kw] cluster, she produced a slightly labialized [k^{w}], which was due to her own idiolect rather than any dialect characteristic. It is probable that [k^{w}] would be common in all dialects.

An interesting error was that concerning the ford *flicker*, which was heard by only 58.73% of the listeners. 15.87% heard *liquor*, easily explainable by the fact that *flicker* is preceded by *group*; [p] and [f] are quite similar and [f] might easily be mistaken for the aspiration of [p]. But 23.81% of the listeners heard *quicker*. Even if it is assumed that the [p] creates confusion in the following word, there is no basis for explaining the perception of [k^{w}] where [fl] was produced. In the alternate response form of this item, when the speaker pronounced *quicker*, 100.00% responded correctly. It can only be assumed that *flicker* is a word with high confusion tendencies, because of the low intensity of the [fl] cluster.

71.43% of the listeners heard *rage* correctly. The remaining listeners responded randomly among the other choices; four listeners did not respond at all.

Anger was mistaken for *anchor* 23.81% of the time; as in Speaker J.H.'s lists, voicing is confused, a function of the test words rather than dialect.

The last case of confusion in the utterances of the Columbus speakers is between *confer* and *confirm*. The word immediately following is *verse*; those listeners who heard *confirm* must have overcompensated for voicing, inserting a labial consonant between [r] and [v].

Other errors of these types occur in the responses to speakers of the other dialects. These kinds of errors will not be analyzed as they are functions of the test, and not induced by dialect. However, it should be noted that a greater number of test-induced errors occurred in the other four dialects than in the Columbus dialect, thus suggesting that overall intelligibility is affected by dialect, but not in predictable dialect errors.

One of the most outstanding features of the New York dialects is the distortion, or absence of [r] following a vowel. Many of the confusions shown in the lists of the New York speaker, M.G. (Lists 2 and 3), occurred in words containing [r].

Only 66.67% of the listeners responded correctly to *horror*. 3.17% heard *father* and *borrow*, respectively, and 26.98% heard *power*. The production of *horror* showed a short [ɔ] instead of [o] and the [ə] which nearly always replaces [r] sounded very like a [w]; the semivowel was made necessary for the transition to the next syllable, which was [ʌŋ]. In syllables ending with a vowel followed by a word or syllable beginning with a vowel, as *horror* is pronounced in New York, [r] is often intrusive. However, dissimilating influences prevent the introduction of [r] in this position.[1] [w] is quite a common replacement for [r] in child language; thus it is predictable that listeners who are unfamiliar with a New York [r] would hear [w].

Speaker M.G.'s [r]-sounds tend to resemble [w] in all positions. This peculiarity is not to be considered a functionally defective [r], since it is heard throughout this dialect area. It seems evident that the [r] distortion creates confusion with other liquids, such as [l] and [w], as occurred when the speaker pronounced *grow*. 7.94% heard *glow*, and 9.52% heard *go* with no liquid at all.

When *drift* was pronounced, only 12.70% of the listeners responded correctly. 49.21% heard *drip*, which can be explained in a manner similar to the arguments presented for Speakers C.B. and J.H., but 38.10% heard *thrift*. A [w]-like [r] would have a longer voicing feature than a clear [r] and a [d] with a weak onset might easily be mistaken for a [θ]. It is also common in this dialect for initial dental stops to be slightly affricated.

The responses generated bypproduction of *gull* are nearly random, but explainable by the New York substitution of [ɑ] for [ʌ] in stressed positions. Thus, 74.60% of the listeners heard the back vowel, responding with *gall*, *gold*, or *goal*.

Analysis of Speaker B.N.'s productions (Lists 3 and 4) were made difficult by the high percentage of listeners who did not indicate any responses.

In many cases nearly all listeners who responded did so correctly, but percentage scores in these cases are only between 60% and 80%; as a result, it is impossible to guess what the listeners thought they heard; they could not decide themselves. Therefore, only those items with a significant number of wrong answers indicated will be looked at.

Most of the errors in Speaker B.N.'s dialect are consonant confusions of manner; a few are errors in place of articulation. Speaker B.N. also exhibits diphthongization of vowels, common in the Southern speech area. This tendency has diffused throughout the Kentucky, West Virginia, and Southern Ohio area, creating what might be mistaken for a "Southern drawl." It is probable that this is the cause of amny of the no answer responses.

Two confusions are due to the backness of the [l], which occurs in both the Columbus and the Southern Ohio dialects. 28.57% of the listeners heard *virtual* when *virtue* was pronounced. In both dialects. the two words sound nearly alike; unstressed syllabic [l] often suggests [u] or [o], and the two words are easily confused. In the second case, 11.11% heard *meadow* when *mettle* was pronounced. Medial [d] and [t] are usually flapped, and the syllabic [l] immediately following the flap is articulated so far in the back of the mouth as to suggest [o].

Of the mistakes in manner of articulation, the most consistent is *spear* (34.92%) for *sphere*. *Sphere* is one of only a few English words with an [sf] cluster and would probably be confused in any dialect.

22.22% of the listeners mistook *kernel* for *curdle*, a nasal for its homorganic stop. 44.44% heard *burst* for *birch*; an [st] cluster for a prepalatal affricate, [tʃ]. The word immediately following is *praise* which could suggest a final stop rather than a fricative release.

When *shave* was pronounced, nearly 27% heard *shade*, a dissimilation from [f] in *effect*, the word following. Other confusions of this type occur as do mistakes in place of articulation; many more than occurred in the other tests. It is interesting that most of the listeners laughed when they heard the first few utterances of this speaker--perhaps an indication that they thought this dialect was very different from their own.

One example illustrates the similarity between the vowels [ɛ] and [ɪ] in the two Ohio dialects. When *ten* was pronounced (after *chain*), only 14.29% responded dorrectly. The remaining answers were nearly random between *pen*, *pin* , *tent*, and N.A. Here the stop confusion is not dialect related, but in the alternate response list, pronounced by Speaker M.G., nearly one-third of the listeners mistook *pen* for *pin*; since these two vowels merge in the Ohio dialects, the listeners would only differentiate them with careful listening, if at all.

Final [t] in a cluster is lost in Southern dialects. This is illustrated by this speaker where only 65.08% of the listerners heard *plant*.

The British dialect, spoken by Speaker G.D. (Lists 4 and 5), also shows a number of items with high percentages of N.A. responses, although this tendency was not consistent throughout the test. It was noticed that most listeners tended to have either a great deal of trouble, or little at all with this dialect. Relatively few scores are near the average, but at either end of the scale.

Intervocalic [r] is flapped in this dialect as are [t] and [d] in American dialects, so when *storage* was pronounced, it suggested a medial [t]; 17.46% of the listeners heard *shortage*.

The consonants of this speaker which involve oral pressure at some level seem to be characterized by their firmness, e.g. the onset is somewhat stronger than normal, thus some confusions in voicing result, as between *folly* and *volley*, *smashing* and *matching*. Other consonant confusions were mainly of manner (*reverse* for *revert*), but few of the errors show percentages over 15%. The items where the correct responses were marked less than 85% of the time were usually the items with high percentages of no answer. The extremely clipped quality of this dialect produces only a few test induced assimilation errors. Since most of the errors were not consistent, little else can be said about dialectal influences on the test responses.

Most of the errors indicated in Speaker C.D.'s Black dialect (List 6) are confusions of final consonants and clusters, although there are a few vowel-diphthong mistakes. In both lists for this speaker, *prod* and *proud* was confused, although *proud* was taken for *prod* more often than the reverse situation. The speaker diphthongizes all stressed vowels, and the resulting diphthong is typically similar to the sound of its first element. Thus words like *prod* and *proud* are nearly undistinguishable in this dialect. The tendency is a residual quality of Southern Negro speech which is frequently heard even in the Northern urban areas of the country.

Some consonant confusions occur which are not dialect derived, mostly initial stop confusions and voiced stop-spirant confusions, but wherever the final consonant is the crucial element, mistakes occurred. Black speakers tend to drop or obscure final consonants in general, also a residue of Southern Negro speech; thus errors occurred for: *new*, *noon*, *nude*; *law*, *log*; *term*, *turn*; *flat*, *flak*; *print*, *prince*; *wake*, *wait*, *wade*; *blast*, *black*; *jump*, *junk*.

Tint and *tense* were confused, but besides the problem of final consonants, there is the merging of [ɪ] and [ɛ] which occurs also in the Ohio dialects.

In urban dialects in general, [θ] often goes to [t]. Indications of this occurred in the test when 20.63% of the listeners heard *fateful* when *faithful* was pronounced. Confusion also occurred between *suit* and *shoot*, but this is not believed to have been caused by the dialect of the speaker, but rather by his tendency to distort [s] to a slight degree.

Conclusion

The most intelligible speakers to listeners of the dialect of Columbus, Ohio, are speakers of the Columbus dialect. Relative intelligibility varies with dialect; dialects arranged in order of most to least intelligible are: Columbus, British, Black, New York, and Rural Ohio.

Unfortunately, only a few specific instances of dialect features are extractable from the mass of results for each list. Direct comparisons between lists are only possible for a list and its alternate response list. Some deviations occur in one list which do not occur in its alternate, suggesting differences between speaker-dialect intelligibility, but comparing successive lists is difficult because the test words are different.

A serious problem arose in evaluating the data--that of the test-induced assimilation errors. Although the number of these errors varies from dialect to dialect, they tend to obscure the general results. It is ironic that the reason for which the test was chosen, the phraselike structure of the test items, was the reason that the data were so difficult to interpret. Scoring the tests is quite simple, but the process of extracting frequencies of all responses is very time-consuming, since it must be done by hand.

Therefore, in the opinion of the experimenter, the usefulness of the test as a measure of dialect intelligibility is somewhat overshadowed by the assimilation errors caused by the testing procedure. Although the results did yield predicted variations, some amount of judgment was necessary to determine which errors were test-induced and irrelevant to the purpose of the study. However, it is believed that the multiple-choice format is the most desirable for studies of this kind. The great number of N.A. responses indicates that a greater number of blank spaces would occur in a write-down test for naive listeners because they simply would not know what to write down.

Footnote

[1] Horror is seldom pronounced correctly by speakers of any dialect. What is usually heard is /hor·/.

Bibliography

Black, John W. "Multiple-Choice Intelligibility Tests," Journal of Speech and Hearing Disorders 22, 213-235. 1957.

Black, John W. Multiple-Choice Intelligibility Test. Interstate Printers and Publishers, Inc., Danville, Ill. 1963.

Clarke, Frank R. Technique for Evaluation of Speech Systems. Stanford Research Institute, Menlo Park. Calif. 1965.

Harms, L. S. "Listener Comprehension of Speakers of Three Status Groups," Language and Speech 4, 109-112. 1961.

Miller, G. A., and P. S. Nicely, "An Analysis of Perceptual Confusions Among Some English Consonants," Journal of the Acoustical Society of America 27, 338-352. 1965.

Reed, Carroll, Dialects of American English. World Publishing Co., Cleveland. 1967.

Wise, Claude M. Applied Phonetics. Prentice-Hall, Inc., Englewood Cliffs., N.J. 1957.

LIST #1

RESPONSE	C.B.	J.H.	RESPONSE	C.B.	J.H.
FORM	00.00	0.00	GROUP	98.41	0.00
WARM	0.00	100.00	TROOP	0.00	0.00
SWARM	100.00	0.00	COUPE	0.00	0.00
STORM	0.00	0.00	FRUIT	1.59	100.00
N.A.	0.00	0.00	N.A.	0.00	0.00
CAMPUS	4.76	98.41	QUICKER	23.81	100.00
CANVAS	95.24	1.59	FLICKER	58.73	0.00
PAMPHLET	0.00	0.00	SLICKER	1.59	0.00
PANTHER	0.00	0.00	LIQUOR	15.87	0.00
N.A.	0.00	0.00	N.A.	0.00	0.00
COURT	42.86	4.76	BEEF	80.95	0.00
FORT	0.00	0.00	BEAST	4.76	0.00
PORT	7.94	95.24	BEAT	12.70	0.00
QUART	49.21	0.00	BEAM	0.00	100.00
N.A.	0.00	0.00	N.A.	1.59	0.00
AIRFORCE	1.59	0.00	REASON	1.59	0.00
AIRPORT	98.41	0.00	REGION	7.94	1.59
AIRCORPS	0.00	98.41	LEGION	84.13	22.22
AIRBORNE	0.00	1.59	LEGEND	4.76	76.19
N.A.	0.00	0.00	N.A.	1.59	0.00
SPARK	0.00	0.00	WONDER	87.30	0.00
PARK	3.17	0.00	BLUNDER	3.17	100.00
DARK	3.17	98.41	THUNDER	6.35	0.00
BARK	92.06	1.59	SPONSOR	0.00	0.00
N.A.	1.59	0.00	N.A.	3.17	0.00
TASSEL	98.41	1.59	CORN	1.59	0.00
TACKLE	1.59	0.00	TORN	0.00	100.00
CATTLE	0.00	0.00	HORN	96.83	0.00
PASTEL	0.00	98.41	BORN	0.00	0.00
N.A.	0.00	0.00	N.A.	1.59	0.00

RESPONSE	C.B.	J.H.	RESPONSE	C.B.	J.H.
STRETCH	1.59	0.00	RAID	6.35	93.65
THREAT	90.48	1.59	RATE	6.35	6.35
DREAD	3.17	98.41	RANGE	9.52	0.00
BREAD	0.00	0.00	RAGE	71.43	0.00
N.A.	4.76	0.00	N.A.	6.35	0.00
HEAR	0.00	0.00	FITTING	0.00	100.00
STEER	1.59	0.00	PRETTY	0.00	0.00
NEAR	0.00	100.00	CITY	96.83	0.00
DEER	98.41	0.00	SITTING	0.00	0.00
N.A.	0.00	0.00	N.A.	3.17	0.00
GUARD	1.59	0.00	OWL	1.59	0.00
HEARTEN	1.59	96.83	CALL	0.00	0.00
GARDEN	96.83	1.59	HALL	7.94	98.41
BARGAIN	0.00	0.00	ALL	85.71	1.59
N.A.	0.00	1.59	N.A.	4.76	0.00
CURTAIN	85.71	1.59	UNCLE	6.35	0.00
PERTAIN	0.00	0.00	BUCKLE	1.59	1.59
PERSON	1.59	0.00	KNUCKLE	90.48	98.41
CERTAIN	11.11	98.41	STUCCO	0.00	0.00
N.A.	1.59	0.00	N.A.	1.59	0.00
EXPORT	87.30	0.00	DREAD	0.00	0.00
EXTORT	0.00	98.41	DRESS	96.83	1.59
EXPERT	6.35	0.00	REST	3.17	98.41
ESCORT	1.59	0.00	RED	0.00	0.00
N.A.	4.76	1.59	N.A.	0.00	0.00
FILE	0.00	98.41	SCREECH	84.13	0.00
PANEL	0.00	0.00	PREACH	3.17	0.00
FUNNEL	1.59	0.00	REACH	3.17	0.00
FINAL	95.24	1.59	STREET	7.94	100.00
N.A.	3.17	0.00	N.A.	1.59	0.00

LIST #2

RESPONSE	M.G.	C.B.	RESPONSE	M.G.	C.B.
SKID	100.00	12.70	HEART	76.19	98.41
SKIN	0.00	0.00	BARGE	0.00	0.00
HID	0.00	85.71	LARD	0.00	0.00
HIT	0.00	1.59	HARD	25.40	1.59
N.A.	0.00	0.00	N.A.	1.59	0.00
MOVE	68.25	3.17	FASTEN	85.71	1.59
MOOD	33.33	1.59	PASSION	3.17	3.17
FOOD	0.00	92.06	FASHION	7.94	0.00
SMOOTH	0.00	0.00	PASSING	1.59	95.24
N.A.	0.00	3.17	N.A.	3.17	0.00
SWIM	0.00	1.59	ANGLE	1.59	0.00
TWIN	0.00	95.24	AMBER	1.59	0.00
SWIFT	0.00	0.00	ANGER	93.65	23.81
TWIST	100.00	1.59	ANCHOR	3.17	76.19
N.A.	0.00	1.59	N.A.	1.59	0.00
PROCLAIM	12.70	0.00	YOKE	1.59	96.83
DOMAIN	0.00	100.00	JOKE	98.41	3.17
COCAINE	0.00	0.00	CHOKE	1.59	0.00
PROFANE	88.89	0.00	DOPE	0.00	0.00
N.A.	0.00	0.00	N.A.	0.00	0.00
SPIN	7.94	0.00	CHAT	3.17	96.83
PIN	6.35	96.83	CHAP	6.35	1.59
THIN	69.84	1.59	SHACK	28.57	0.00
FIN	15.87	1.59	SHAFT	63.49	1.59
N.A.	1.59	0.00	N.A.	0.00	0.00
REPEAT	0.00	1.59	HEADING	0.00	0.00
RECEIVE	95.24	0.00	SITTING	0.00	96.83
RECEDE	6.35	0.00	KNITTING	100.00	1.59
REPRIEVE	0.00	96.83	FITTING	0.00	0.00
N.A.	0.00	1.59	N.A.	0.00	1.59

RESPONSE	M.G.	C.B.	RESPONSE	M.G.	C.B.
COURT	4.76	7.94	PIPE	74.60	3.17
CORD	0.00	92.06	PIKE	25.40	0.00
HORSE	0.00	0.00	TYPE	0.00	95.24
COURSE	95.24	0.00	TIGHT	0.00	1.59
N.A.	0.00	0.00	N.A.	0.00	0.00
BALANCE	92.06	1.59	BEAST	92.06	0.00
BALLOT	4.76	96.83	BEAT	3.17	98.41
GALLONS	0.00	0.00	MEAT	0.00	1.59
VALID	4.76	0.00	LEAST	4.76	0.00
N.A.	0.00	1.59	N.A.	0.00	0.00
DRANK	11.11	1.59	DRAY	0.00	0.00
RANK	88.89	0.00	GREY	1.59	1.59
RANCH	0.00	0.00	SPRAY	96.83	0.00
DRAG	0.00	96.83	PRAY	1.59	98.41
N.A.	0.00	1.59	N.A.	0.00	0.00
BANKING	0.00	0.00	THRIFT	38.10	1.59
FLANKING	3.17	98.41	DRIP	49.21	0.00
LANKY	96.83	0.00	DRIFT	12.70	0.00
BLANKET	0.00	0.00	GRIP	0.00	98.41
N.A.	0.00	1.59	N.A.	0.00	0.00
BORROW	3.17	93.65	CONFIRM	19.05	19.05
HORROR	66.67	1.59	CONFER	1.59	80.95
FATHER	3.17	0.00	CONSERVE	20.63	0.00
POWER	26.98	3.17	CONCERN	57.14	0.00
N.A.	0.00	1.59	N.A.	1.59	0.00
UNFOLD	88.89	3.17	VERSE	7.94	92.06
UNTOLD	6.35	3.17	FIRST	87.30	4.76
CONTROLLED	0.00	0.00	BURST	4.76	3.17
UPHOLD	4.76	92.06	HURT	0.00	0.00
N.A.	0.00	3.17	N.A.	0.00	0.00

LIST #3

RESPONSE	B.N.	M.G.	RESPONSE	B.N.	M.G.
DEED	0.00	1.59	DIMMER	6.35	0.00
WEED	3.17	0.00	DINNER	84.13	0.00
SEED	6.35	95.24	THINNER	1.59	98.41
FEED	69.84	0.00	TINNER	0.00	1.59
N.A.	20.63	3.17	N.A.	7.94	1.59
PROTRUDE	1.59	3.17	ENVY	52.38	0.00
CONCLUDE	73.02	3.17	EMPTY	1.59	0.00
CONSTRUED	1.59	92.06	ENTRY	20.63	100.00
INCLUDE	4.76	0.00	ENDING	0.00	0.00
N.A.	19.05	1.59	N.A.	23.81	0.00
TRAIN	68.25	1.59	RUMOR	69.84	12.70
CRANE	9.52	1.59	ROAMER	12.70	85.71
STRAIN	1.59	7.94	RUBBER	0.00	0.00
TERRAIN	3.17	87.30	ROVER	1.59	1.59
N.A.	17.46	1.59	N.A.	15.87	0.00
VIRTUAL	28.57	96.83	SPHERE	55.56	0.00
CURFEW	3.17	0.00	FEAR	3.17	1.59
VIRTUE	61.90	3.17	SPEAR	34.92	0.00
VIRGIN	0.00	0.00	BEER	0.00	96.83
N.A.	7.94	0.00	N.A.	6.35	1.59
HIDE	3.17	0.00	GULL	7.94	25.40
FIVE	0.00	0.00	GALL	6.35	12.70
HIRE	80.95	0.00	GOLD	63.49	39.68
FIRE	1.59	100.00	GOAL	15.87	30.16
N.A.	12.70	0.00	N.A.	6.35	0.00
PACK	3.17	0.00	PETAL	4.76	0.00
PATCH	85.71	0.00	METTLE	77.78	1.59
CATCH	4.76	3.17	MEADOW	11.11	0.00
CAT	0.00	96.83	SETTLE	0.00	98.41
N.A.	6.35	0.00	N.A.	6.35	0.00

RESPONSE	B.N.	M.G.	RESPONSE	B.N.	M.G.
FAULT	74.60	7.94	GLOW	6.35	7.94
VAULT	12.70	85.71	GO	90.48	9.52
DOG	0.00	0.00	GROW	0.00	82.54
FOG	0.00	1.59	GOAT	0.00	0.00
N.A.	12.70	4.76	N.A.	3.17	0.00
BURST	44.44	74.60	LATE	3.17	100.00
HURT	9.52	3.17	LADEN	1.59	0.00
FIRST	6.35	12.70	LAZY	0.00	0.00
BIRCH	23.81	4.76	LADY	92.06	0.00
N.A.	15.87	1.59	N.A.	3.17	0.00
TRADE	3.17	4.76	BREAK	80.95	66.67
TRACE	6.35	95.24	RAKE	7.94	14.29
PRAISE	71.43	0.00	GREAT	3.17	15.87
PRAY	4.76	0.00	GRAPE	3.17	3.17
N.A.	14.29	0.00	N.A	4.76	0.00
BLACK	3.17	1.59	CHANGE	34.92	9.52
TRACK	0.00	98.41	CHAIN	50.97	49.21
SLACK	90,48	0.00	STAIN	1.59	1.59
FLAK	1.59	0.00	SHAME	1.59	39.68
N.A.	4.76	0.00	N.A.	12.70	0.00
KERNEL	22.22	0.00	PEN	26.98	30.16
CURDLE	61.90	6.35	PIN	17.46	66.67
TURTLE	11.11	1.59	TENT	25.40	1.59
HURDLE	0.00	92.06	TEN	14.29	1.59
N.A.	4.76	0.00	N.A.	15.87	0.00
GRAFT	0.00	3.17	HARD	12.70	0.00
DRAFT	6.35	68.25	PART	17.46	0.00
DRAB	63.49	28.57	HARSH	3.17	98.41
GRAB	26.98	0.00	HEART	53.97	0.00
N.A.	3.17	0.00	N.A.	12.70	0.00

LIST #4

RESPONSE	G.D.	B.N.	RESPONSE	G.D.	B.N.
STARDOM	3.17	0.00	EIGHT	93.65	0.00
PARDON	84.13	1.59	ACHE	3.17	1.59
GARDEN	0.00	98.41	HATE	0.00	96.83
AUTUMN	1.59	0.00	BAKE	0.00	0.00
N.A.	11.11	0.00	N.A.	3.17	1.59
CALL	1.59	0.00	REVOLVE	0.00	7.94
BALL	6.35	96.83	INVOLVE	0.00	0.00
HALL	79.37	0.00	RESOLVE	1.59	88.89
SMALL	1.59	0.00	DISSOLVE	95.24	0.00
N.A.	12.70	3.17	N.A.	3.17	3.17
BUBBLE	7.94	0.00	NEEDLE	95.24	3.17
STUBBLE	1.59	93.65	FETAL	0.00	3.17
TROUBLE	4.76	1.59	EAGLE	1.59	0.00
DOUBLE	76.19	3.17	BEETLE	0.00	88.89
N.A.	9.52	1.59	N.A.	3.17	4.76
TOP	88.89	3.17	ABLE	0.00	0.00
HOP	0.00	0.00	STABLE	0.00	0.00
POP	7.94	9.52	FABLE	93.65	1.59
PROP	1.59	87.30	TABLE	1.59	92.06
N.A.	1.59	0.00	N.A.	4.76	4.76
TOOL	1.59	88.89	RECLINE	88.89	9.52
CRUEL	92.06	6.35	REFINE	4.76	6.35
DROOL	1.59	1.59	RECLAIM	3.17	4.76
COOL	1.59	0.00	REPLY	0.00	73.02
N.A.	3.17	3.17	N.A.	3.17	6.35
STORAGE	76.19	6.35	FOLLY	12.70	73.02
PORRIDGE	0.00	87.30	VOLLEY	82.54	19.05
SHORTAGE	17.46	4.76	POLISH	0.00	0.00
STORY	3.17	0.00	TROLLEY	0.00	0.00
N.A.	3.17	1.59	N.A.	4.76	4.76

RESPONSE	G.D.	B.N.	RESPONSE	G.D.	B.N.
GAVE	0.00	0.00	CLAD	3.17	3.17
SHADE	92.06	26.98	CLAN	9.52	6.35
FADE	3.17	0.00	PLAN	79.37	12.70
SHAVE	1.59	68.25	PLANT	1.59	65.08
N.A.	1.59	4.76	N.A.	4.76	12.70
EFFECT	9.52	77.78	LIFT	88.89	31.75
EXPECT	0.00	0.00	RIFT	3.17	14.29
INSPECT	3.17	1.59	DRIFT	3.17	12.70
INFECT	84.13	9.52	LIST	1.59	23.81
N.A.	3.17	11.11	N.A.	3.17	17.46
HARD	1.59	84.13	BEHAVE	1.59	0.00
CARD	92.06	4.76	WITHHOLD	6.35	9.52
CORD	1.59	1.59	REVOLT	0.00	73.02
HARSH	0.00	1.59	BEHOLD	88.89	3.17
N.A.	4.76	7.94	N.A.	3.17	14.29
STRANGE	19.05	0.00	QUARRY	0.00	9.52
BRING	11.11	0.00	GLORY	92.06	33.33
RAIN	3.17	88.89	GORY	3.17	53.97
BRAIN	58.73	1.59	SORRY	0.00	0.00
N.A.	9.52	9.52	N.A.	0.00	4.76
WAD	1.59	77.78	SUCH	1.59	73.02
WASH	1.59	4.76	TOUCH	1.59	7.94
SQUAD	79.37	4.76	NUT	96.83	1.59
SQUASH	9.52	1.59	BUTT	0.00	6.35
N.A.	7.94	11.11	N.A.	0.00	11.11
PLANT	3.17	0.00	FORCE	100.00	7.94
CLAMP	4.76	4.76	FOURTH	0.00	6.35
CRAMP	15.87	85.71	COURSE	0.00	3.17
TRAMP	69.84	0.00	HORSE	0.00	76.19
N.A.	6.35	9.52	N.A.	0.00	6.35

LIST #5

RESPONSE	J.H.	G.D.	RESPONSE	J.H.	G.D.
COOK	12.70	1.59	TOOK	0.00	90.48
CROOK	87.30	0.00	SHOOK	93.65	0.00
BROOK	0.00	98.41	SHOCK	6.35	1.59
BOOK	0.00	0.00	COCK	0.00	0.00
N.A.	0.00	0.00	N.A.	0.00	7.94
FAIR	95.24	1.59	OPEN	0.00	1.59
BARE	3.17	6.35	OBOE	4.76	65.08
CARE	0.00	1.59	OPAL	93.65	20.63
PAIR	0.00	90.48	OVAL	1.59	3.17
N.A.	0.00	0.00	N.A.	0.00	9.52
ANNUAL	0.00	0.00	TRIAL	15.87	4.76
AMPLE	11.11	1.59	FILE	0.00	0.00
AMBLE	87.30	1.59	FRAIL	3.17	77.78
APPLE	1.59	96.83	TRAIL	82.54	7.94
N.A.	0.00	0.00	N.A.	0.00	7.94
BRINK	12.70	87.30	FLAME	100.00	1.59
BRIDGE	1.59	0.00	BLAME	0.00	1.59
BRISK	0.00	1.59	CLAIM	0.00	95.24
BRICK	84.13	4.76	PLANE	0.00	0.00
N.A.	1.59	4.76	N.A.	0.00	1.59
SKIM	0.00	88.89	WORM	7.94	3.17
HYMN	0.00	3.17	WORK	0.00	1.59
VIM	0.00	0.00	WORD	9.52	92.06
DIM	98.41	0.00	WERE	79.37	3.17
N.A.	1.59	9.52	N.A.	3.17	0.00
ACTION	0.00	0.00	RELIEVE	19.05	71.43
MATCHING	95.24	6.35	RECEIVE	0.00	0.00
MAGIC	3.17	3.17	RELIEF	79.37	28.57
SMASHING	0.00	80.95	RELEASE	1.59	0.00
N.A.	1.59	9.52	N.A.	0.00	0.00

RESPONSE	J.H.	G.D.	RESPONSE	J.H.	G.D.
CLOCK	4.76	100.00	WORLD	60.32	6.35
BLOCK	1.59	0.00	WHIRL	39.68	1.59
PLOT	84.13	0.00	WOOL	0.00	6.35
BLOT	4.76	0.00	WOULD	0.00	84.13
N.A.	4.76	0.00	N.A.	0.00	1.59
KIND	80.95	0.00	HAPPY	0.00	0.00
PINE	9.52	0.00	HANDY	100.00	0.00
FINE	1.59	100.00	CANDY	0.00	96.83
TIME	4.76	0.00	ENVY	0.00	1.59
N.A.	3.17	0.00	N.A.	0.00	1.59
LEAPING	0.00	1.59	DODGE	0.00	96.83
SLEEPING	98.41	0.00	DARK	3.17	0.00
CREEPING	0.00	0.00	DOT	90.48	3.17
REAPING	0.00	98.41	DOCK	4.76	0.00
N.A.	1.59	0.00	N.A.	1.59	0.00
EIGHTY	98.41	1.59	CONSCRIPT	0.00	3.17
ACHING	0.00	0.00	CONFLICT	0.00	0.00
DAINTY	0.00	87.30	ASSIST	0.00	95.24
BABY	1.59	3.17	UNFIT	98.41	0.00
N.A.	0.00	7.94	N.A.	1.59	1.59
PROOF	0.00	87.30	REFER	0.00	1.59
HOOP	0.00	4.76	REHEARSE	6.35	3.17
GROUP	0.00	0.00	REVERSE	93.65	22.22
SWOOP	100.00	0.00	REVERT	0.00	71.43
N.A.	0.00	7.94	N.A.	0.00	1.59
WHIP	0.00	0.00	BUDGET	98.41	0.00
QUIT	84.13	0.00	BUCKET	1.59	98.41
QUICK	15.87	1.59	BUNION	0.00	0.00
TWIST	0.00	93.65	BUDGE	0.00	0.00
N.A.	0.00	4.76	N.A.	0.00	1.59

LIST #6

REHEARSE	C.D.	C.D.
SQUIRM	0.00	0.00
FIRM	0.00	0.00
TERM	88.89	38.10
TURN	11.11	61.90
N.A.	0.00	0.00
HATE	96.83	1.59
HASTE	0.00	0.00
EIGHT	3.17	0.00
TAKE	0.00	98.41
N.A.	0.00	0.00
COMMIT	98.41	12.70
SUBMIT	0.00	0.00
PERMIT	0.00	1.59
COMMENCE	1.59	85.71
N.A.	0.00	0.00
CLOUD	0.00	0.00
CROWD	7.94	3.17
PROUD	80.95	39.68
PROD	11.11	57.14
N.A.	0.00	0.00
WAIST	96.83	3.17
WAKE	0.00	50.79
WADE	3.17	25.40
WAIT	0.00	19.05
N.A.	0.00	1.59
FEELING	0.00	6.35
MEETING	7.94	4.76
FEEDING	0.00	87.30
MEANING	92.06	1.59
N.A.	0.00	0.00

REHEARSE	C.D.	C.D.
NEGLECT	0.00	0.00
DEFLECT	95.24	0.00
REFLECT	4.76	98.41
REFLEX	1.59	1.59
N.A.	0.00	0.00
LOST	0.00	0.00
LONG	1.59	0.00
LOG	28.57	57.14
LAW	60.32	42.86
N.A.	9.52	0.00
ROBBER	0.00	98.41
JOBBER	93.65	1.59
HARBOR	3.17	0.00
SHOPPER	3.17	0.00
N.A.	0.00	0.00
HELD	0.00	95.24
BELL	3.17	1.59
FELL	9.52	3.17
TELL	85.71	0.00
N.A.	1.59	0.00
INVITE	88.89	3.17
INSIGHT	6.35	0.00
INSIDE	0.00	6.35
ADVICE	1.59	87.30
N.A.	3.17	3.17
BLAST	0.00	68.25
FLAT	73.02	6.35
FLAK	23.81	6.35
BLACK	1.59	15.87
N.A.	1.59	3.17

RESPONSE	C.D.	C.D.	RESPONSE	C.D.	C.D.
PLAYFUL	0.00	100.00	EGG	3.17	100.00
FAITHFUL	79.37	0.00	EDGE	95.24	0.00
FATEFUL	20.63	0.00	HEDGE	1.59	0.00
BASEBALL	0.00	0.00	HEAD	0.00	0.00
N.A.	0.00	0.00	N.A.	0.00	0.00
SUIT	77.78	0.00	FINDING	0.00	0.00
SHOOT	22.28	0.00	BINDING	100.00	95.24
BOOT	0.00	1.59	BLINDING	0.00	4.76
FRUIT	0.00	98.41	LANDING	0.00	0.00
N.A.	0.00	0.00	N.A.	0.00	0.00
DEPEND	0.00	1.59	TINT	0.00	84.13
DETAIN	15.87	96.83	PRINT	28.57	1.59
BECAME	82.54	0.00	PRINCE	69.84	0.00
RETAIN	1.59	1.59	TENSE	1.59	14.29
N.A.	0.00	0.00	N.A.	0.00	0.00
PLURAL	0.00	0.00	DESK	95.24	3.17
NEUTRAL	0.00	0.00	DECK	0.00	25.24
RURAL	80.95	4.76	DEATH	3.17	0.00
RULER	19.05	95.24	DEBT	0.00	1.59
N.A.	0.00	0.00	N.A.	1.59	0.00
NOUN	4.76	0.00	BOTH	1.59	0.00
NEW	36.51	61.90	BOAT	34.92	100.00
NUDE	15.81	38.10	VOTE	63.49	0.00
NOON	42.86	0.00	QUOTE	0.00	0.00
N.A.	0.00	0.00	N.A.	0.00	0.00
BRAVE	1.59	0.00	YAWN	0.00	0.00
STAVE	6.35	92.06	JUMP	0.00	82.54
BATHE	1.59	1.59	JUNK	0.00	17.46
SAVE	90.48	6.35	YOUNG	100.00	0.00
N.A.	0.00	0.00	N.A.	0.00	0.00

APPENDIX 1

BASED ON 63 LISTENERS

NUMBER WRONG	PERCENTAGE	NUMBER WRONG	PERCENTAGE
1	1.59	33	52.38
2	3.17	34	53.97
3	4.76	35	55.56
4	6.35	36	57.14
5	7.94	37	58.73
6	9.52	38	60.32
7	11.11	39	61.90
8	12.70	40	63.49
9	14.29	41	65.08
10	15.87	42	66.67
11	17.46	43	68.25
12	19.05	44	69.84
13	20.63	45	71.43
14	22.22	46	73.02
15	23.81	47	74.60
16	25.40	48	76.19
17	26.98	49	77.78
18	28.57	50	79.37
19	30.16	51	80.95
20	31.75	52	82.54
21	33.33	53	84.13
22	34.92	54	85.71
23	36.51	55	87.30
24	38.10	56	88.89
25	39.68	57	90.48
26	41.27	58	92.06
27	42.86	59	93.65
28	44.44	60	95.24
29	46.03	61	96.83
30	47.62	62	98.41
31	49.21	63	100.00
32	50.79		

Intensity and Duration Analysis of Hungarian Secondary Stress

Richard Gregorski and Andrew Kerek

Intensity and Duration Analysis of Hungarian Secondary Stress

Richard Gregorski, Ohio State University
Andrew Kerek, Miami (Ohio) University

It is generally agreed that in Hungarian, primary stress always falls on the first syllable of a word. Fónagy (1966) found no consistent acoustic correlate to this stress, but did find a correspondence between the activity of the internal intercostal muscles and stress. However, Magdics' study (1969) seems to indicate that stressed vowels are generally more intense, longer, and higher in pitch than their unstressed counterparts.

The status of secondary stress--both its placement and rhythmic function--has been much disputed (Rákos, 1966). There are two main proposals regarding the placement of secondary stress: position and syllable-length theories.[1] Kerek (in press) attempts to resolve the issue by offering an alternative which accounts for secondary stress placement in terms of context, that is, "on the basis of the speaker's (subconscious) anticipation of the stress conditions in the immediately following context." Closely connected with this theory are certain constraints related to syllable length and unstressed syllable sequences. Despite the general interest in Hungarian secondary stress, there exists, to our knowledge, no experimental research into either its acoustic or physiological basis. It was the purpose of this study to determine to what degree intensity and duration function as acoustic correlates of this secondary stress.

It was assumed that the appearance of secondary stress on a vowel in terms of intensity and duration would manifest itself as an increase of these parameters over the vowel's unstressed counterpart, and not necessarily as absolute intensity or duration prominences over adjacent syllables. This is consistent with the view that stress is correlated with effort of production, i.e., that both stress production and perception involve a knowledge of the intrinsic physical parameters of a syllable and the consequent adjustment of effort needed to mark the presence of stress. Also important in stress analysis is the magnitude of the increase, for it is doubtful that a non-perceivable increment can have any functional significance. It was decided that the general perceptual threshold of ±1 dB for intensity and 10-40 msec. for duration (Lehiste, 1970) would serve as a fair indicator of the potential perceptual significance of intensity and duration increases.

The following set of sentences was chosen for the experiment (´ = primary stress; ˆ = secondary stress):

1. A. [fɛ́ʃtɛt:e:k pɛ́tit] "They painted Pete."
   B. [fɛ́ʃtɛt:ê:k pɛ́tit] "They painted Pete."
2. A. [fɛ́ʃtɛt:ê:tɛk pɛ́tit] "You (pl.) painted Pete."
   B. [fɛ́ʃtɛt:ê:tɛk pɛ́tit] "You (pl.) painted Pete."

2. C. [fɛ́ʃtɛt:e:tɛ̂tit] "You (pl.) painted Pete."
3. A. [fɛ́ʃtɛgɛt:ê:tɛk pɛ́tit] "You (pl.) kept painting Pete."
B. [fɛ́ʃtɛgɛt:ê:tɛk pɛ́tit] "You (pl.) kept painting Pete."
C. [fɛ́ʃtɛgɛ̂t:e:tɛ̂k pɛ́tit] "You (pl.) kept painting Pete."
4. A. [fɛ́ʃtɛgɛ̂thɛt:ê:tɛk pɛ́tit] "You (pl.) may have kept painting Pete."
B. [fɛ́ʃtɛgɛ̂thɛt:ê:tɛk pɛ́tit] "You (pl.) may have kept painting Pete."
5. A. [fɛ́ʃtɛgɛ̂thɛt:ê:tɛk iʃ pɛ́tit] "You (pl.) may have also kept painting Pete."
B. [fɛ́ʃtɛgɛ̂thɛt:ê:tɛk îʃ pɛ́tit] "You (pl.) may have also kept painting Pete."

These sentences were chosen for the following reasons: (1) the numerous voiceless fricatives and plosives would facilitate segmentation; (2) for the most part, the vowel qualities could be kept constant throughout the expanding sequences; and (3) a variety of secondary stress placements could be employed.

The subject (AK), a trained linguist, is a native of Budapest, Hungary, who has lived in the United States since 1957. He constructed the test sentences, which exhibited possible secondary stress patterns in his dialect. He was presented with a randomized list consisting of ten occurrences of each of the sentence patterns (except 2.C. and 3.C.) and was asked to produce the sentences at his normal rate of speech. He was then instructed to produce 2.C. and 3.C. (the alternate secondary stress assignments for 2.B. and 3.B. respectively) ten times each. This procedure was followed since a randomization of 2.C. and 3.C. within the first list might have introduced an uncontrolled variable into the experiment, that is, the subject could have inadvertently substituted 2.C. for 2.B. and 3.C. for 3.B. or vice versa. He then repeated the first list and the alternate patterns. Two additional similar sessions followed at intervals of about a week, at the end of which about 60 productions of each pattern or approximately 720 utterances for the total set had been recorded.

The recorded utterances were processed by a Frøkjaer-Jensen intensity meter and pitch meter, the output of which was converted by an Elema-Schönander Mingograph (100 mm/sec) into a three-channel display: (1) oscillogram, (2) intensity curve, and (3) fundamental frequency pattern. The duration of the vowels was measured to the nearest 1/2 millimeter (i.e., 5 milliseconds). The intensity of the vowels was measured in terms of peak sound pressure level in dB relative to an arbitrary level.

Table I presents the intensity results. There were no differences between the vowels with secondary stress and their unstressed counterparts. Note that there was a 1 dB difference between the unstressed [ɛ]'s of -[tɛt]- of 2.A-C and between the unstressed [ɛ]'s of -[tɛt]- of 4.A-B. However, these differences did not occur between similar unstressed vowels within the other sentences.

TABLE I
AVERAGE INTENSITY OF VOWELS IN UTTERANCES OF VARIOUS LENGTHS (in dB)
(Secondary stressed vowels underlined)

Sentence Type	Syllable Type					
	tε(t)	gεt	hεt	tə:(k)	tεk	iȷ
1A	43			41		
1B	43			$\underline{41}$		
2A	43			$\underline{41}$	42	
2B	44			$\underline{41}$	42	
2C	43			41	$\underline{42}$	
3A	43	43		$\underline{41}$	41	
3B	43	43		$\underline{41}$	41	
3C	43	$\underline{43}$		41	$\underline{41}$	
4A	43	$\underline{43}$	42	$\underline{41}$	41	
4B	43	$\underline{43}$	42	$\underline{41}$	42	
5A	43	$\underline{43}$	42	$\underline{41}$	41	38
5B	43	$\underline{43}$	42	$\underline{41}$	41	$\underline{38}$

Table II presents the duration results. There was a 1-7 msec. difference between unstressed vowels of the same syllable sequence with the A-B-C comparisons and also between the secondary stressed vowels of the same syllable sequences in the A-B-C comparisons. In six of the seven unstressed versus secondary stressed comparisons, the unstressed vowel was longer than its secondary stressed counterpart; the range of these differences was 6-12 msec. In only one comparison (1A-b) was the secondary stressed vowel longer; the difference was 14 msec.

TABLE II
AVERAGE DURATION OF VOWELS IN UTTERANCES OF VARIOUS LENGTHS (in msec.)
(Secondary stressed vowels underlined)

Sentence Type	Syllable Type					
	tɛ(t)	gɛt	hɛt	te·(k)	tɛk	ij
1A	72			74		
1B	67			88		
2A	71			88	70	
2B	72			83	66	
2C	66			76	56	
3A	58	74		87	69	
3B	58	74		80	64	
3C	51	67		68	55	
4A	56	80	59	89	70	
4A	56	79	55	83	64	
5A	55	81	54	86	68	57
5B	56	80	54	84	65	51

Since the average differences fall below the just noticeable differences, intensity and duration cannot be considered as acoustic correlates of secondary stress. However, since the fundamental frequency of the vowel comparisons had not been analyzed, this parameter could not be ruled out as a possible correlate. To determine if this was a promising direction for a future study, a perceptual test was given to the subject to see if indeed he could perceive the stress patterns that he had produced. The subject was presented with a tape of twenty randomized productions of the sentences:

2. B. [fɛ́jtɛt:ê:tɛk pɛ́tit] "You (pl.) painted Pete."
   C. [fɛ́jtɛt:e:tɛ̂k pɛ́tit] "You (pl.) painted Pete."

and twenty randomized productions of the sentences:

3. B. [fɛ́jtɛgɛ̭t:ê:tɛk pɛ́tit] "You (pl.) kept painting Pete."
   C. [fɛ́jtɛgɛ̀t:e:tɛ̂k pɛ́tit] "You (pl.) kept painting Pete."

These were the two sets of sentences in which alternate secondary stress assignments occurred. The subject was asked to assign secondary stress to each sequence. He correctly identified 6 out of 20 sequences in the 2.B-C set, and 10 out of 20 sequences in the 3.B-C set. Hence, his judgments were random. We conclude that an explanation of Hungarian secondary stress in terms of acoustic and perceptual correlates does not seem promising.

Footnote

[1] Most linguists who have commented on Hungarian stress hold that secondary stress occurs on the third and every subsequent odd-numbered syllable of a word, i.e. according to numerical syllable position. Some linguists, notably Szinnyei and Lotz, point out that a short third (and any odd-numbered) syllable causes the stress to shift to the following even-numbered syllable; hence, in this view, the relevant condition is the length value of a syllable. For references, see Kerek (in press).

Bibliography

Bánhidi, Zoltán, Zoltán Jókay, and Dénes Szabó. Learn Hungarian, Second Edition. Budapest: Tankönyvkiadó. 1965.

Fónagy, Iván. "Electro-physiological and Acoustic Correlates of Stress and Stress Perception." Journal of Speech and Hearing Research 9, 231-244. 1966.

Kerek, Andrew. Hungarian Metrics: Some Linguistic Aspects of Iambic Verse. Indiana University Publications, Uralic and Altaic Series, Vol. 117. Bloomington, Indiana. In press.

Lehiste, Ilse, and G. E. Peterson. "Vowel Amplitude and Phonemic Stress in American English." Journal of the Acoustical Society of America 31, 428-435. 1959.

Lehiste, Ilse. Suprasegmentals. Cambridge, Mass.: The M.I.T. Press. 1970.

Magdics, Klara. Studies in the Acoustic Characteristics of Hungarian Speech Sounds. Indiana University Publications, Uralic and Altaic Series, Vol. 97. Bloomington, Indiana. 1969.

Nemser, William J. and Francis S. Juhasz. A Contrastive Analysis of Hungarian and English Phonology. American Council of Learned Societies, Research and Studies in Uralic and Altaic Languages, Project No. 70. 1964.

Peterson, G. E. and Ilse Lehiste. "Duration of Syllable Nuclei in English." Journal of the Acoustical Society of America 32, 693-703. 1960.

Rákos, Petr. Rhythm and Metre in Hungarian Verse. Praha: Universita Karlova. 1966.

Tarnóczy, Tamás. "Can the Problem of Automatic Speech Recognition be Solved by Analysis Alone?" Rapports du 5e Congrés International d'Acoustique, Vol. II. Conférences generales. Liége: D. D. Commins, pp. 371-387. 1965.

Experiments with Synthetic Speech Concerning Quantity in Estonian

Ilse Lehiste

Experiments with Synthetic Speech Concerning Quantity in Estonian

Ilse Lehiste

1. Introduction

This paper constitutes a first report on an experiment designed to test the relevance of various suprasegmental parameters in the perception of quantity in Estonian. The test materials consisted of synthetically produced acoustic stimuli, intended to sample systematically the acoustic spaces containing the minimal triples taba - tapa - tappa and sada - saada! - saada. The synthesis was performed by means of a Digital Data Processor (DDP 24) computer at the Bell Telephone Laboratories.[1] The synthesis was carried through entirely by rule, i.e. no attempt was made to imitate a known speaker. The stimuli will be described below in more detail. Test tapes containing randomized stimuli were presented to 26 listeners, who are native speakers of Estonian, at the Experimental Phonetics Laboratory of the Academy of Sciences in Tallinn, Estonia.[2] Two tapes were used, one for the taba - tapa - tappa set, the other for the sada - saada! - saada set; each contained 252 stimuli. As there were 26 listeners and each made 504 judgments, the data consist of 13,104 individual judgments. The statistical evaluation of the materials is in progress; however, some results are already available, and a preliminary survey is given below.

2. Taba - tapa - tappa

The synthetic material was designed to test the ranges of /p/ durations which would be assigned to the three quantities, and the contribution of second syllable duration to the perception of the three test words. The duration of /p/ was varied in twenty-one 10 msec steps over a continuous range from 40 to 240 msec. Each of the 21 /p/-durations was combined with three durations for the second vowel: 180 msec, 120 msec, and 90 msec. The duration of the first vowel was kept constant at 120 msec; the fundamental frequency was likewise constant (at 120 Hz). The total of 21 x 3 = 63 stimuli was arranged in four different randomizations and presented to listeners, who had to assign each stimulus to one of the three words taba, tapa or tappa. The listeners thus made a forced-choice linguistic judgment rather than a phonetic judgment. Each listener gave 252 responses, for a total of 6,552 responses. The results of the listening test are summarized on the following figures and tables.

Table 1 and Figure 1 show the general effect of second syllable duration on the assignment of the words to quantities one, two and three. It is obvious that a second syllable duration of 180 msec favors assignment to quantities one and two: the number of taba and tapa responses is greatest under this condition. On the other hand, a second syllable duration of 90 msec favors assignment of the word to quantity three.

Tables 2-4 and Figures 2-4 show the number of judgments as taba, tapa or tappa as a function of the duration of intervocalic /p/. Each of the three tables and figures represents judgments associated with one of the three second syllable durations. The discussion of the tables and the figures will be limited to a few brief comments.

If we consider the crossing-points of curves representing taba, tapa, and tappa judgments as 'phoneme boundaries' between quantities 1, 2 and 3 of the intervocalic consonant, then we note that the phoneme boundary between /p/ in quantity 1 and /p/ in quantity 2 depends only slightly on the duration of the second vowel: with decreasing second syllable duration, the boundary shifts from approximately 110 msec for a second syllable duration of 180 msec to 105 msec for a second syllable of 120 msec, and to 100 msec for a second syllable of 90 msec. However, the boundary between quantities 2 and 3 appears crucially affected by the duration of the second syllable. Figure 2 shows that if the second syllable had a duration of 180 msec, the boundary between tapa and tappa was at 225 msec, and even with the longest duration, 240 msec, the differentiation between long /p/ and overlong /p/ was very tenuous. With second syllables of 120 and 90 msec, the boundary between long and overlong intervocalic /p/ occurred at 175 and 170 msec respectively.

3. Sada - saada! - saada

The set of test items designed to test the perception of quantity in disyllabic words of the type sada - saada! - saada is a little more complicated. This time there were three variables: duration of the vowel of the first syllable, duration of the vowel of the second syllable, and the fundamental frequency pattern distributed over the two syllables. The duration of the first vowel varied in seven 20-msec steps from 120 to 240 msec, while the duration of intervocalic /t/ was kept constant at 60 msec. Each of the first syllables was combined with the same three second syllable durations as in the previous case, namely 180 msec, 120 msec, and 90 msec. Furthermore, each disyllabic stimulus was synthesized with three fundamental frequency patterns: a level pattern (monotone at 120 Hz), a step-down pattern (with the first syllable level at 120 Hz and the second syllable level at 80 Hz), and a falling pattern (first syllable falling from 120 Hz to 80 Hz, second syllable level at 80 Hz). The total number of stimuli was again 7 x 3 x 3 = 63, the total number of items on the randomized tape was 252, and the number of judgments was 6,552.

The results are presented on Tables 5-8 and Figures 5-11. Again, only a few descriptive comments will be given this time.

Table 5 and Figure 5 show the influence of second syllable duration and fundamental frequency pattern on the overall classification of stimuli as sada, saada! and saada. As is apparent from the left half of Figure 5, the influence of second syllable duration was comparable to what was observed with the set taba - tapa - tappa: a longer second syllable favored judgments for quantities 1 and 2, and disfavored judgments as quantity 3, while the shortest second syllable increased the number of quantity 3 judgments in a substantial manner.

This effect is, however, rather limited compared to the influence of the fundamental frequency pattern. As becomes apparent from Figure 5, the monotone condition was relatively neutral. The step-down pattern, with the first syllable level at 120 Hz and the second syllable level at 80 Hz, produced the greatest number of quantity 2 judgments and the smallest number of quantity 3 judgments. It is important here to notice that the step-down pattern actually decreased quantity 1 judgments; for quantity 1, the monotone pattern was the most favorable one.

Conversely, the falling pattern significantly increased the number of quantity 3 judgments and decreased quantity 2 judgments. This decrease took place almost exclusively at the expense of quantity 2, since the number of quantity 1 judgments remained practically constant.

The phoneme boundaries for the duration of the first vowel are rather difficult to establish, since both the second syllable duration and especially the fundamental frequency pattern have such a strong influence on perception. Some of the problems are illustrated on the figures.

Figure 6 shows the assignment of stimuli to quantities 1, 2 and 3 with a second syllable of 180 msec and with a level fundamental frequency pattern. It may be recalled that these two conditions favor assignments to quantity 1 and disfavor assignments to quantity 3. As is obvious from the figure, the overlap between quantities 1 and 2 occurs at approximately 160 msec, while the two curves representing quantities 2 and 3 do not overlap at all. Even at the longest duration, 240 msec, 73 out of 104 judgments were still made in favor of quantity 2.

Figure 7 shows the number of judgments with the same second syllable duration--180 msec--but with a falling fundamental frequency pattern on the first syllable. As was mentioned before, this pattern favors assignments to quantity 3 and disfavors assignments to quantity 2, leaving quantity 1 practically unaffected. The phoneme boundary between quantities 1 and 2 has shifted only very slightly, from 160 msec to approximately 155 msec. It is now also possible to talk about a phoneme boundary between quantities 2 and 3: it would fall at about 210 msec.

Figure 8 shows assignments to the three quantities with a short second syllable (90 msec) and monotone fundamental frequency. As may be remembered, the short second syllable favors assignments to quantity 3, while the monotone fundamental frequency pattern is relatively neutral. A characteristic of all three curves is the extensive overlap between them and the fact that all three curves peak at approximately 75%. The reliability of recognition here obviously was not very great; the phoneme boundaries, however, seem not to have been affected.

Figure 9 shows assignments to the three quantities under conditions maximally favoring quantity 3: a short second syllable (90 msec) and a falling fundamental frequency pattern. The reduction of the number of quantity 2 judgments is particularly striking: even at the 160 msec duration, which produced the greatest number of quantity 2 judgments, their number did not exceed 64 (out of 104). The phoneme boundary between quantities 1 and 2 is not affected, but the boundary between quantities 2 and 3 has now shifted from 210 to 175 msec. The peak of the curve has shifted from 180 msec with level fundamental frequency (Figure 8) to 160 msec.

Figures 10 and 11 summarize the influence of fundamental frequency patterns on assignment to quantities 2 and 3. The second syllable in these two sets of examples was constant at the most neutral, intermediate value, namely at 120 msec.

Figure 10 shows assignments to quantity 2. It is obvious that the left-hand slope of the curve depends very little on the fundamental frequency pattern: the phoneme boundary between quantities 1 and 2 is barely affected by the fundamental frequency. On the other hand, the position of the peak and the phoneme boundary of quantity 2 with regard to quantity 3 are both strongly affected: the peak shifts from about 210 msec with the step-down curve to 180 for the monotone and to 160 for the falling pattern.

The converse situation appears on Figure 11, which shows the influence of fundamental frequency on assignments to quantity 3. Here the neutral pattern produced the smallest number of assignments, the step-down pattern increased the number of quantity 3 judgments somewhat (although the curve never reached 70%), and the falling pattern both steepened the slope of the curve and made it reach a higher peak. It should be noted that even with the falling fundamental frequency pattern the highest number of quantity 3 judgments was 90 out of 104. The peak value for quantity 3 judgments for the whole set of conditions was reached when both conditions were met: the fundamental frequency had a falling pattern and the second syllable was short.

Let me now summarize briefly where we stand with regard to the status of the experiments. I am currently in the process of working out the statistical design for testing the significance of the relationships displayed on this set of tables and figures. I intend to compute correlations between the variables and the judgments and establish the relative contribution of each variable. Until this part of the project is completed, the results are somewhat impressionistic. Nevertheless, it is possible to draw some tentative generalizations.

First of all, I think it is clear that the assignment of a word to a quantity depends not only on the duration of a first syllable vowel or an intervocalic consonant, but also on the duration of the second syllable and on the fundamental frequency pattern applied to the word as a whole. If one defines the point of overlap between two distribution curves as the boundary between two phonemic quantities, one may claim that the placement of these boundaries depends significantly on both second syllable duration and fundamental frequency. I believe that this observation lends support to the notion that what we are dealing with is a higher-level suprasegmental pattern distributed over the whole disyllabic word, not with independently functioning segmental quantity.

It is interesting, furthermore, that the boundary between quantities 2 and 3 is more strongly affected by the pattern applied to the word as a whole than the boundary between quantities 1 and 2. In a very tentative sense, one might find support here for the idea that the older two-way opposition between short and long is more firmly segmentally anchored than the relatively new three-way opposition between short, long and overlong. The older opposition is mainly segmental; the newer three-way opposition is mainly based on differences between patterns manifested over the whole disyllabic word. The implications of these results will become clearer when the statistical analysis is complete.

Footnotes

[1] The DDP 24 computer is a machine of medium size (12K) and speed (5 microseconds). The synthesis programs were written by B. E. Caspers (B. E. Caspers, "Software Facilities and Operating System of a DDP- 224 Computer", Bell Telephone Laboratories, Murray Hill, N.J., 1968). I am grateful to Dr. P. B. Denes, Head of the Speech and Communication Research Department, Bell Telephone Laboratories, for his assistance.

[2] I am indebted to Mr. Kullo Vende for his invaluable help in arranging for the listening sessions. I would also like to thank all individuals who participated in the listening tests.

Table 1. Judgments depending on second syllable duration.

Duration of V_2 in msec	taba	tapa	tappa	Total
180	784	1090	310	2184
120	686	767	731	2184
90	656	731	797	2184
Total	2126	2588	1838	6552

Table 2. Judgments depending on the duration of /p/

V_2 = 180 msec

Duration of /p/ in msec	taba	tapa	tappa
40	104		
50	103		1
60	104		
70	104		
80	103	1	
90	97	7	
100	78	26	
110	50	54	
120	26	78	
130	9	93	2
140	3	100	1
150	2	102	
160		98	6
170		93	11
180		92	12
190	1	80	23
200		71	33
210		61	43
220		56	48
230		45	59
240		33	71
Total	784	1090	310

Table 3. Judgments depending on the duration of /p/

V_2 = 120 msec

Duration of /p/ in msec	taba	tapa	tappa
40	104		
50	103	1	
60	102	2	
70	99	5	
80	96	8	
90	83	21	
100	58	45	1
110	33	71	
120	4	97	3
130	1	98	5
140	3	97	4
150		82	22
160		81	23
170		73	31
180		31	73
190		27	77
200		14	90
210		8	96
220		3	101
230		3	101
240			104
Total	686	767	731

Table 4. Judgments depending on the duration of /p/

V_2 = 90 msec

Duration of /p/ in msec	**taba**	**tapa**	**tappa**
40	104		
50	102	2	
60	103	1	
70	100	4	
80	89	15	
90	67	35	2
100	53	51	
110	31	72	1
120	5	91	8
130		97	7
140		98	6
150		84	20
160	1	76	27
170		50	54
180		23	81
190	1	22	81
200		11	93
210		5	99
220		1	103
230			104
240		1	103
Total	656	731	797

Table 5

Judgments depending on second syllable duration (fundamental frequency patterns combined)

Duration of V_2 in msec	sada	saada!	saada	Total
180	717	1114	353	2184
120	596	1054	534	2184
90	569	942	673	2184
Total	1882	3110	1560	6552

Judgments depending on fundamental frequency pattern (second syllable durations combined)

F_0 pattern (in Hz)	sada	saada!	saada	Total
120-120/120	669	1096	419	2184
120-120/80	605	1326	253	2184
120-80/80	608	688	888	2184
Total	1882	3110	1560	6552

Table 6. Judgments depending on first syllable duration and fundamental frequency pattern (second syllable duration constant at 180 msec)

F_o pattern (in Hz)	V_1 duration (in msec)	sada	saada!	saada	Total
120-120/120	120	101	3		
	140	89	15		
	160	52	51	1	
	180	17	84	3	
	200	1	93	10	
	220		87	17	
	240		73	31	
Total		260	406	62	728
120-120/80	120	96	8		
	140	85	16	3	
	160	42	57	5	
	180	10	84	10	
	200	3	94	7	
	220		89	15	
	240	1	75	28	
Total		237	423	68	728
120-80/80	120	99	5		
	140	72	31	1	
	160	41	58	5	
	180	5	78	21	
	200	2	60	42	
	220	1	45	58	
	240		8	96	
Total		220	285	223	728
		717	1114	353	2184

Table 7. Judgments depending on first syllable duration and fundamental frequency pattern (second syllable duration constant at 120 msec)

F_o pattern (in Hz)	V_1 duration (in msec)	sada	saada!	saada	Total
120-120/120	120	95	8	1	
	140	77	27		
	160	23	72	9	
	180	10	82	12	
	200	2	77	25	
	220	1	61	42	
	240	1	34	69	
Total		209	361	158	728
120-120/80	120	96	8		
	140	78	25	1	
	160	17	83	4	
	180	7	90	7	
	200		92	12	
	220		92	12	
	240	1	70	33	
Total		199	460	69	728
120-80/80	120	87	15	2	
	140	69	33	2	
	160	17	75	12	
	180	10	58	36	
	200	1	27	76	
	220	3	12	89	
	240	1	13	90	
Total		188	233	307	728
		596	1054	534	2184

Table 8. Judgments depending on first syllable duration and fundamental frequency pattern (second syllable duration constant at 90 msec)

F_0 pattern (in Hz)	V_1 duration (in msec)	sada	saada!	saada	Total
120-120/120	120	74	26	4	
	140	76	27	1	
	160	32	63	9	
	180	14	77	13	
	200	3	68	33	
	220	1	40	63	
	240		28	76	
Total		200	329	199	728
120-120/80	120	78	25	1	
	140	58	44	2	
	160	22	78	4	
	180	9	86	9	
	200		87	17	
	220	1	69	34	
	240	1	54	49	
Total		169	443	116	728
120-80/80	120	87	17		
	140	79	19	6	
	160	15	64	25	
	180	14	37	53	
	200	1	20	83	
	220	2	8	94	
	240	2	5	97	
Total		200	170	358	728
		569	942	673	2184

Figure 1. Number of judgments as taba, tapa or tappa, expressed as a function of the duration of the second syllable.

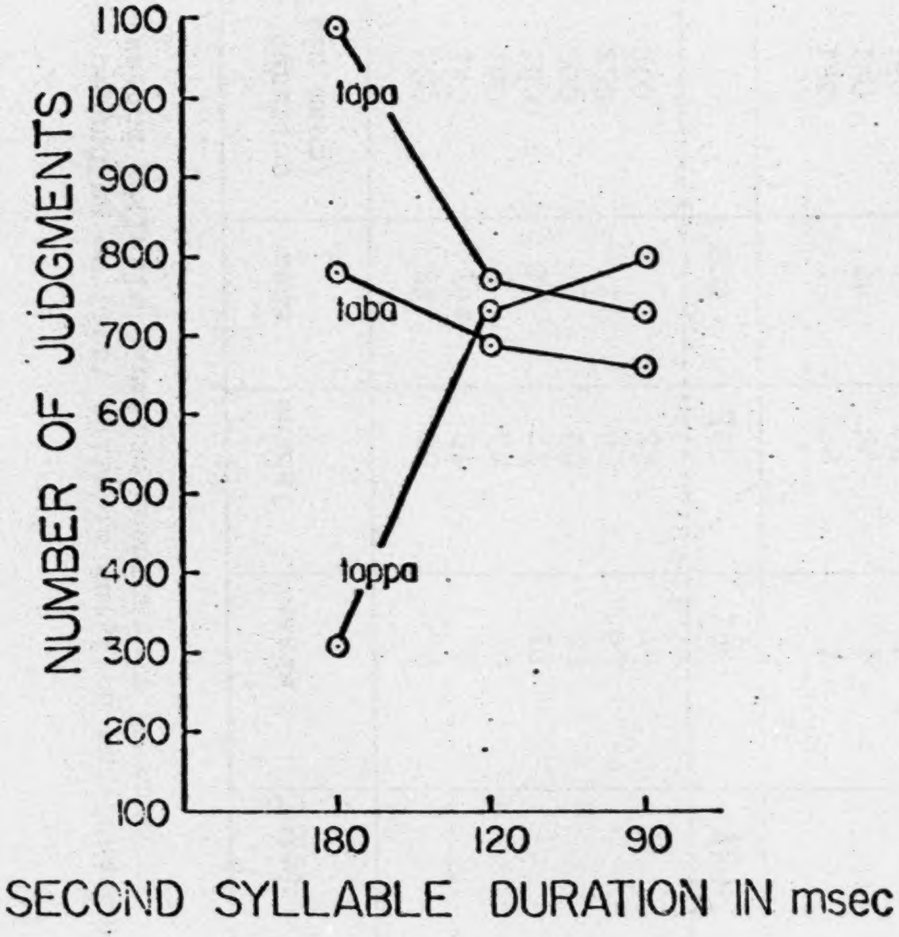


Figure 2. Number of judgments as taba, tapa or tappa, expressed as a function of the duration of intervocalic /p/. Duration of the second syllable was constant at 180 msec.

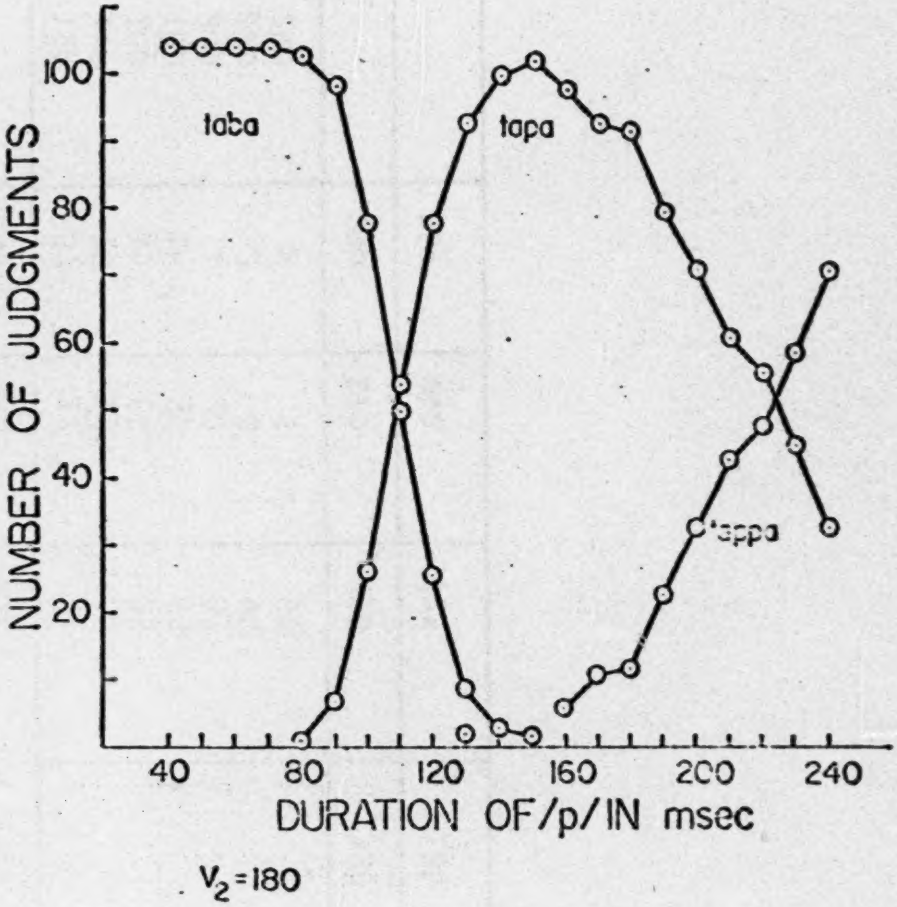


Figure 3. Number of judgments as taba, tapa or tappa, expressed as a function of the duration of intervocalic /p/. Duration of the second syllable was constant at 120 msec.

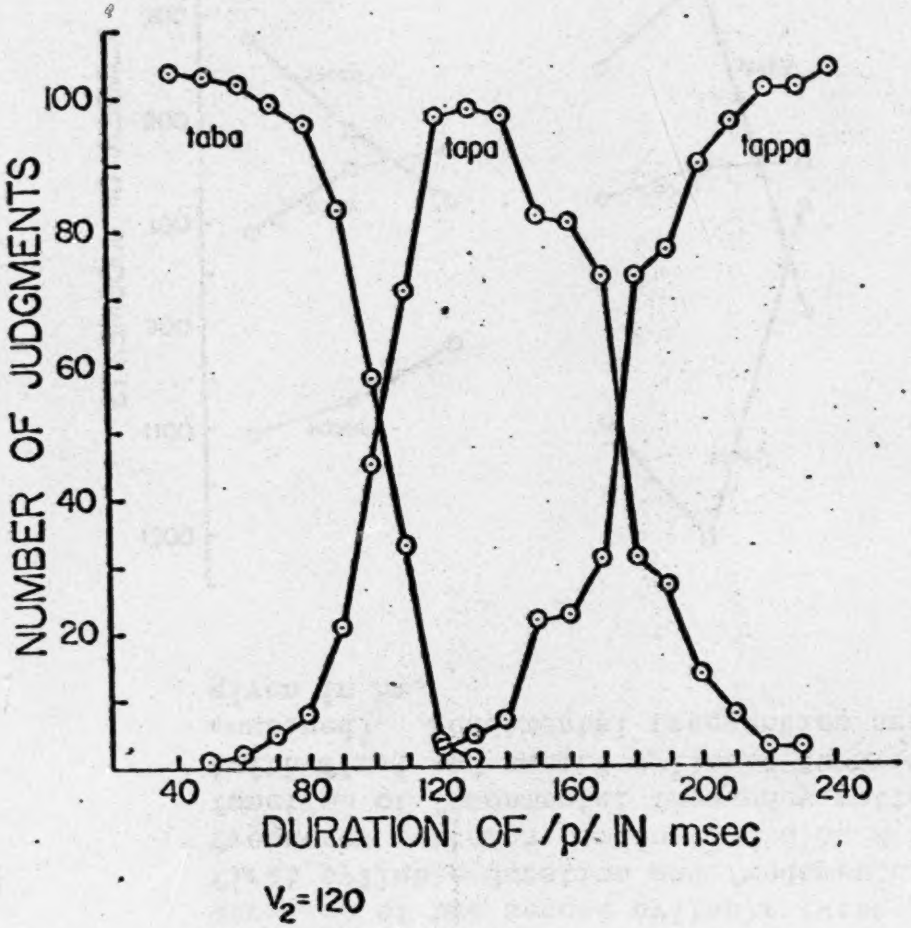


Figure 4. Number of judgments as taba, tapa or tappa, expressed as a function of the duration of intervocalic /p/. Duration of the second syllable was constant at 90 msec.

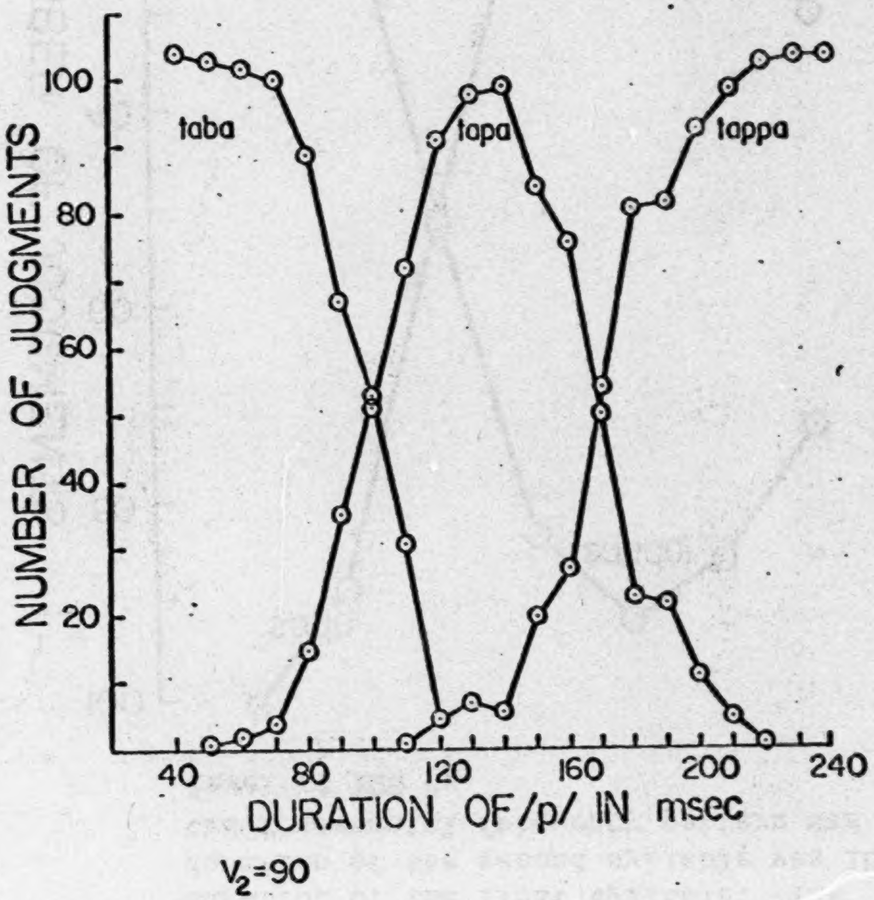


Figure 5. Number of judgments as *sada*, *saada!* or *saada*, expressed as a function of the duration of the second syllable (with first syllable duration and fundamental frequency patterns combined) and as a function of fundamental frequency pattern (with first and second syllable durations combined). Fundamental frequencies are given in Hz.

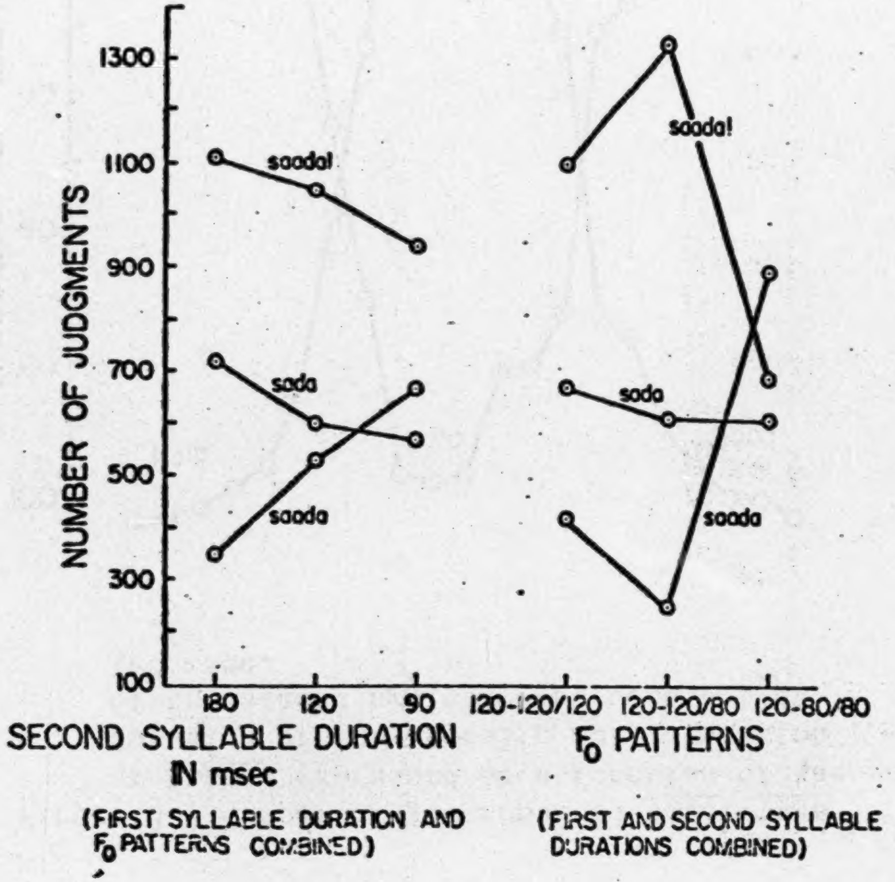


Figure 6. Number of judgments as *sada*, *saada!* or *saada*, expressed as a function of the duration of the first syllable. The duration of the second syllable was 180 msec, the fundamental frequency pattern was level at 120 Hz.

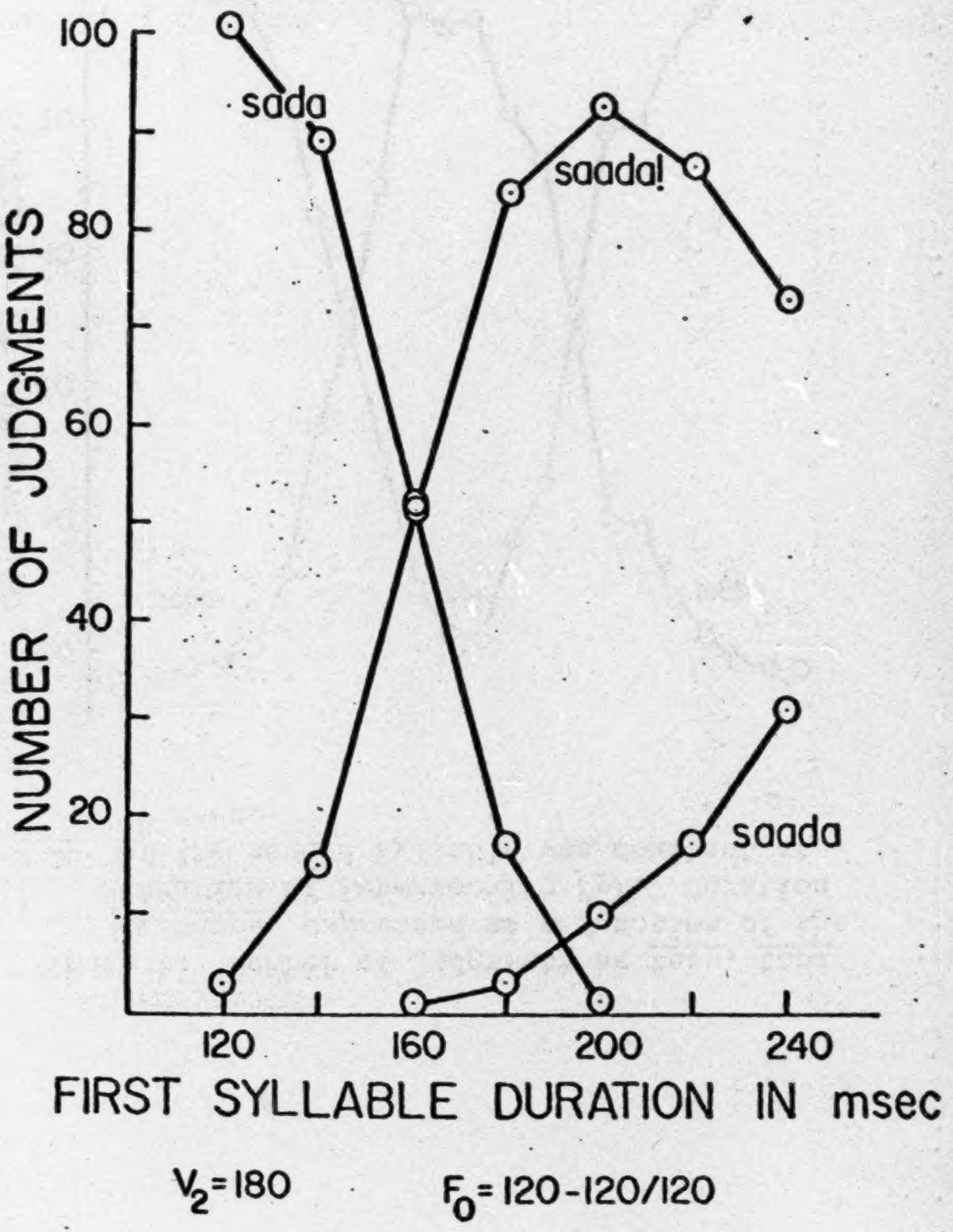


Figure 7. Number of judgments as sada, saada! saada, expressed as a function of the duration of the first syllable. The duration of the second syllable was 180 msec, the fundamental frequency pattern was falling during the first syllable.

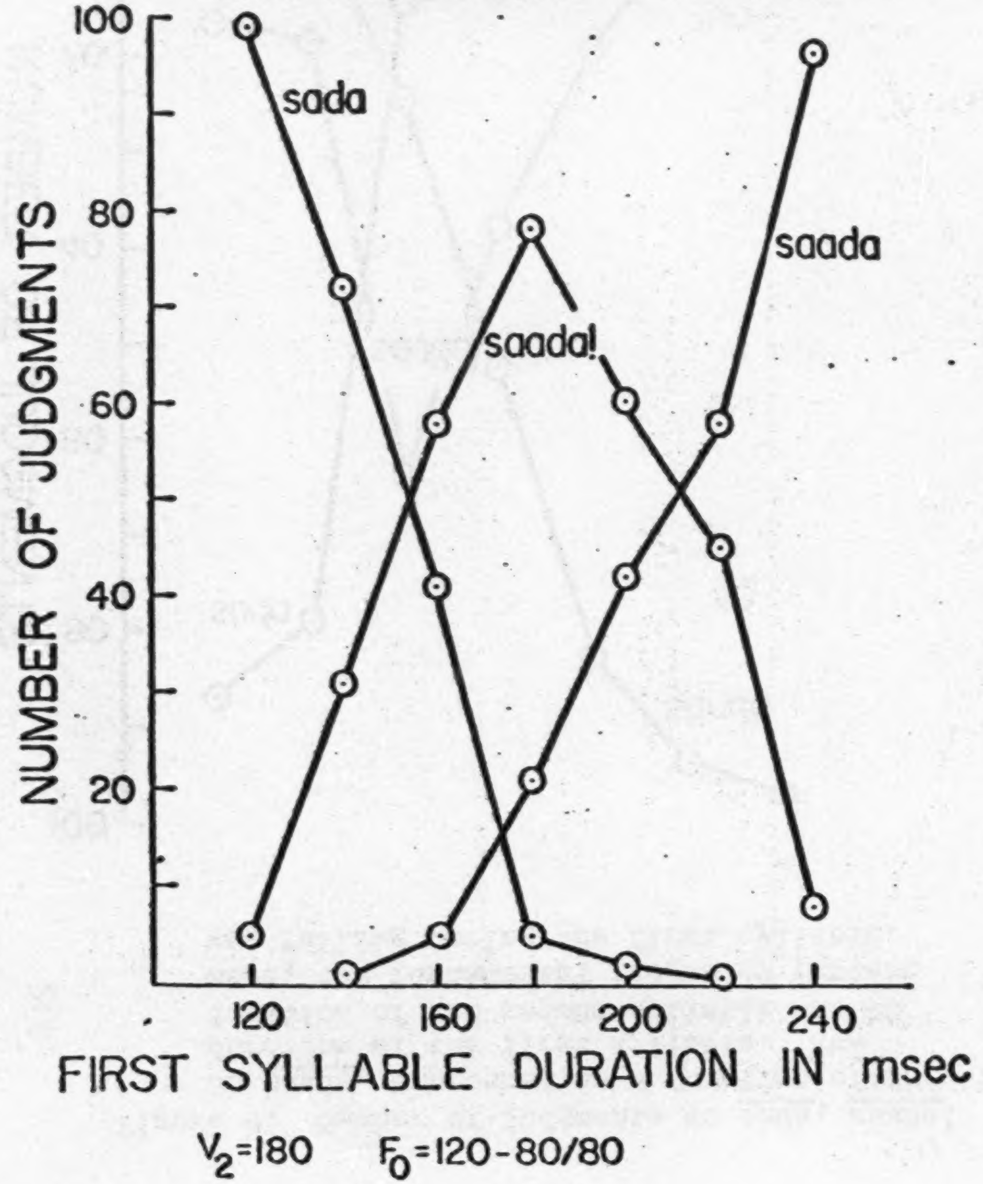


Figure 8. Number of judgments as sada, saada! or saada, expressed as a function of the duration of the first syllable. The duration of the second syllable was 90 msec, the fundamental frequency pattern was level at 120 Hz.

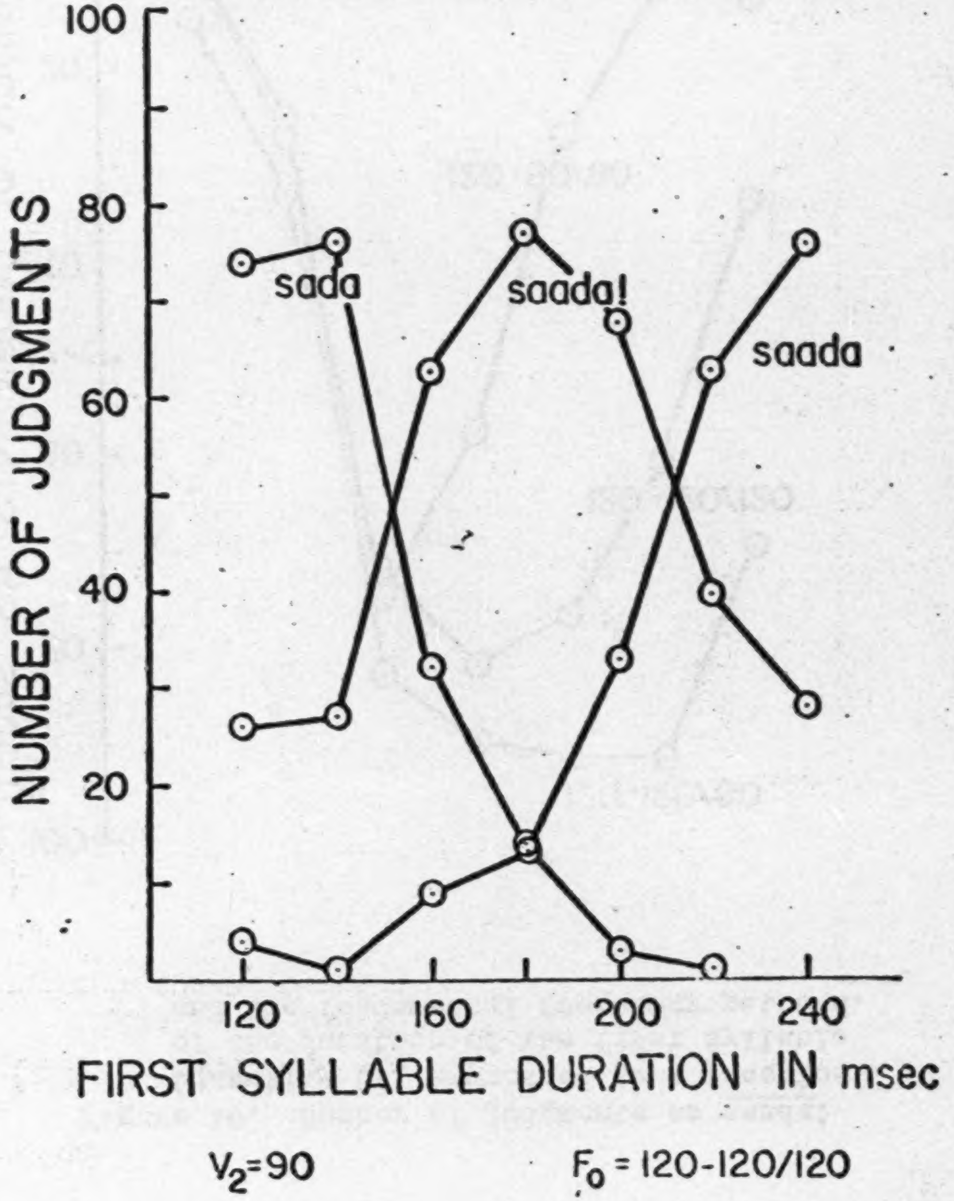


Figure 9. Number of judgments as sada, saada! or saada, expressed as a function of the duration of the first syllable. The duration of the second syllable was 90 msec, the fundamental frequency pattern was falling during the first syllable.

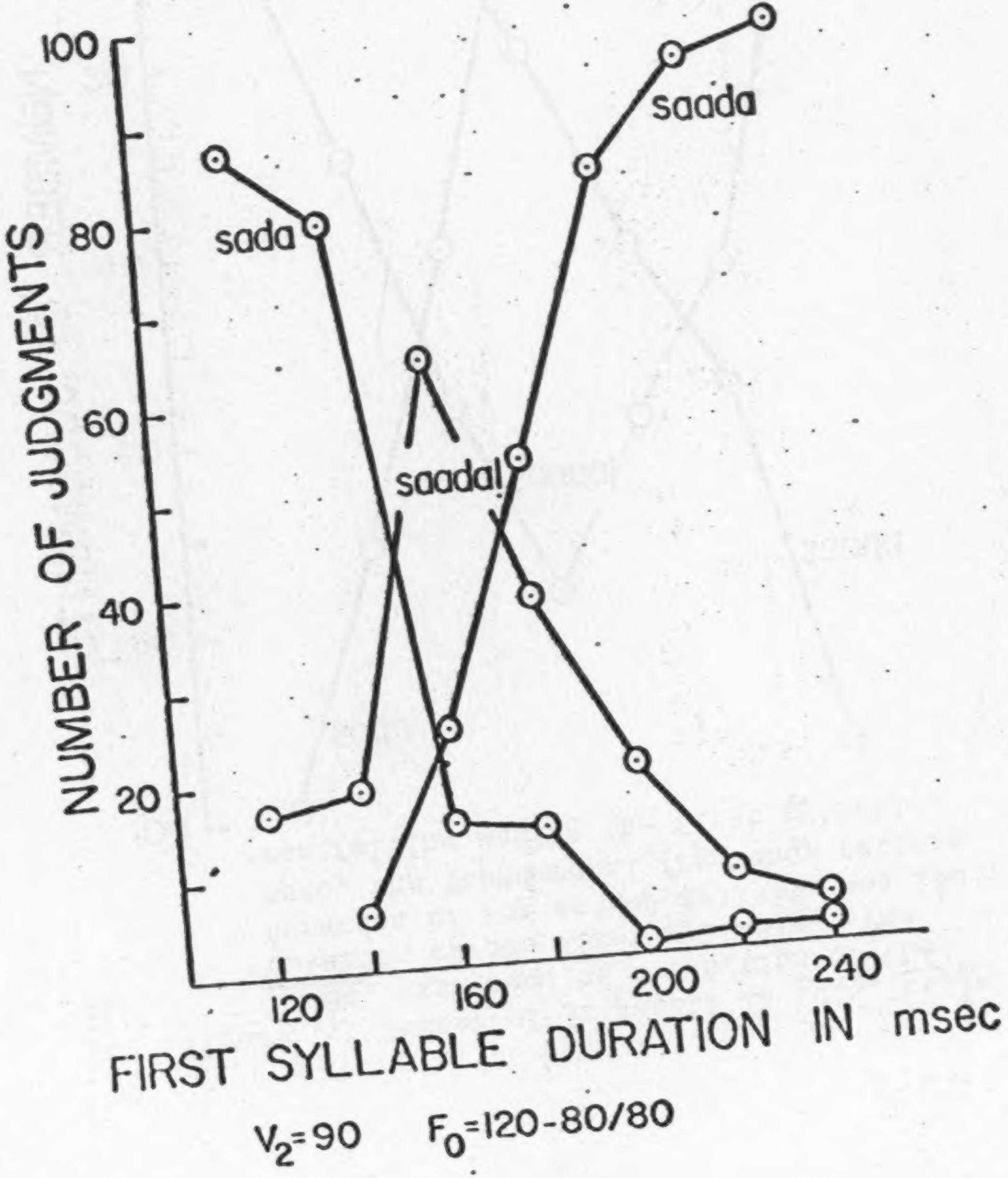


Figure 10. Number of judgments as saada! (quantity 2), expressed as a function of the duration of the first syllable and the fundamental frequency pattern.

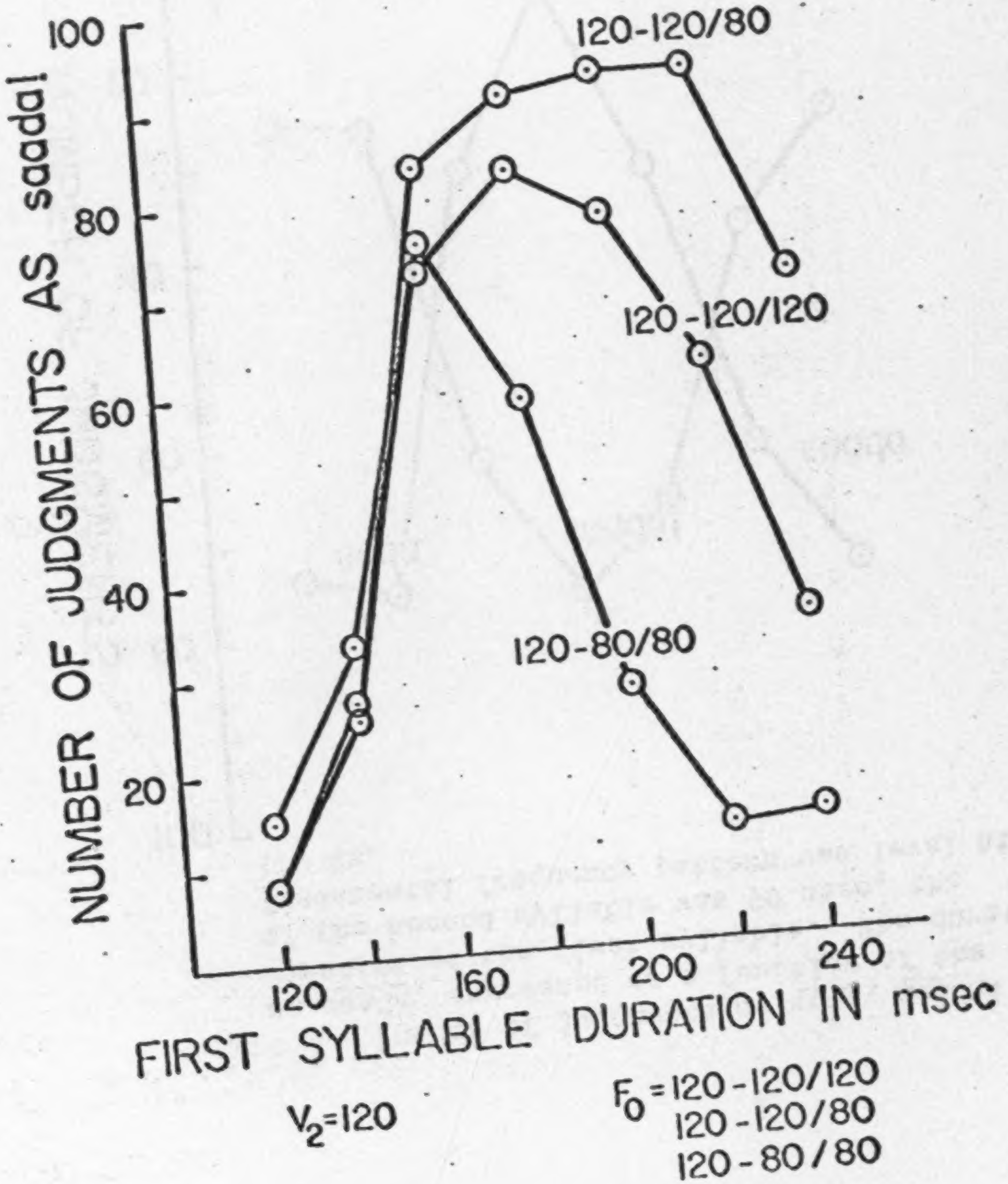
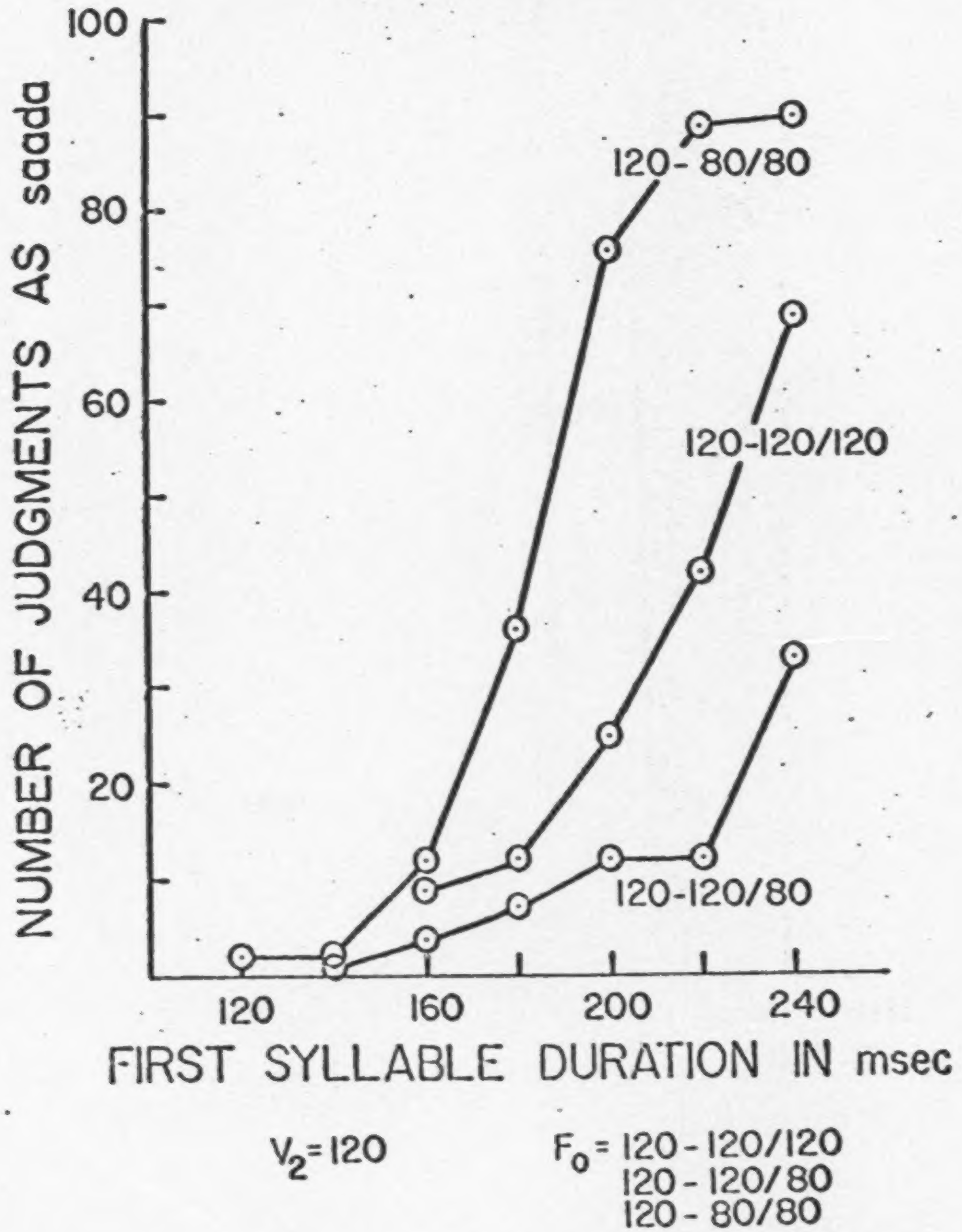


Figure 11. Number of judgments as saada (quantity 3), expressed as a function of the duration of the first syllable and the fundamental frequency pattern.



Phonological Rules in Lithuanian and Latvian

Zinny S. Bond

Phonological Rules in Lithuanian and Latvian*

Zinny S. Bond

Introduction

Lithuanian and Latvian are quite closely related languages, Latvian traditionally being considered the more innovating of the two. The two languages present an ideal case for a comparison of their grammars in terms of shared phonological rules.

In the ideal case, two independently developed grammars of the languages would be compared. However, though there is an extensive treatment of Latvian phonology in a generative framework (Halle and Zeps, 1966), recent work on Lithuanian has been primarily concerned with an analysis of accent. Only Heeschen (1967) has considered other phonological phenomena, and his treatment of Lithuanian phonology is also primarily concerned with accent assignment.

I will simply assume that the analysis of Latvian phonology is basically sound and see which of the Latvian rules are applicable in Lithuanian. If the rules developed for Latvian can also be shown to operate in Lithuanian, then the rules in question can be established as shared by the two languages. The interesting questions in this comparison concern not so much the fact of shared rules, but the place of innovations in the two grammars, as well as changes in the form and applicability of rules.

This paper will be limited to rules primarily involved in the derivation of verbs, though obviously some of the rules are quite general. First, the rules developed by Halle and Zeps for Latvian will be surveyed briefly. Then, each rule will be considered in how (or if) it is applicable to Lithuanian. Some Lithuanian rules will also be discussed. Finally, differences between the two sets of rules will be analyzed.

The Latvian Rules

The fundamental phonological processes have been described by Halle and Zeps (Halle and Zeps, 1966; Zeps, 1970). I will describe the rules they have developed and add, for clarity, a few examples of their application. The notation is informal; examples are given in traditional orthography.

---

*This paper was written in the summer of 1970 while the author held an NDEA Title VI Fellowship.

1. k/c rule

Velar stops are replaced by dental affricates before front vowels:

$$\begin{Bmatrix} k \\ g \end{Bmatrix} \rightarrow \begin{Bmatrix} c \\ dz \end{Bmatrix} \; / \; \underline{\quad} \begin{Bmatrix} i \\ \varepsilon \end{Bmatrix}$$

saku 'I say', sacīšu 'I will say'
ruaka 'hand', ruaciņa 'hand' diminutive

2. i/j rule

The rule defines alternations of long vowels with sequences of vowel plus v or j:

$$\begin{Bmatrix} i \\ u \end{Bmatrix} \rightarrow \begin{Bmatrix} j \\ v \end{Bmatrix} \; / \; \underline{\quad} \; V$$

šūt 'to sow', šuvu 'I sowed'
līt "to rain', lija 'it rained'

3. n/i rule

The sequence vowel plus n becomes the sequence vowel plus i or u:

$$n \rightarrow \begin{Bmatrix} i \; / \; \text{Front vowel} \; \underline{\quad} \begin{Bmatrix} C \\ \# \end{Bmatrix} \\ u \; / \; \text{Back vowel} \; \underline{\quad} \begin{Bmatrix} C \\ \# \end{Bmatrix} \end{Bmatrix}$$

The rule accounts for two types of alternations. First, a long vowel can alternate with a vowel + n sequence, as in dzinu 'I drove', and dzīt 'to drive'. Secondly, the rule provides some of the inputs to the metathesis rule, thereby accounting for alternations of the form pruatu 'I know how', pratu 'I knew how'. In the second case, the -n never appears on the surface.

4. ɛ/e rule

ɛ is raised to e before i or j; any number of vowels in a word will be raised as long as there is no intervening back vowel.

$$\varepsilon \rightarrow e \; / \; \underline{\quad} \, i, \; j$$

ɛcɛ̄tu 'I would harrow', ecēsi 'you will harrow'

5. Metathesis

Except where specifically blocked, metathesis applies unconditionally, to all sequences of the appropriate shape. In spite of the notation, there are only two possible outputs of the metathesis rule, represented as ie and ua in the traditional orthography. The second element of these diphthongs is a mid or low central vowel of rather obscure quality: [ə], [ʌ] or even [ɑ].

$$\begin{Bmatrix} \mathrm{ai} \\ \mathrm{au} \\ \varepsilon\mathrm{i} \\ \varepsilon\mathrm{u} \end{Bmatrix} \rightarrow \begin{Bmatrix} \mathrm{ia} \\ \mathrm{ua} \\ \mathrm{i}\varepsilon \\ \mathrm{u}\varepsilon \end{Bmatrix}$$

skrien 'he runs', skrēja 'he ran'
duad 'he gives', deva 'he gave'

6. Ablaut
ε alternates with i in non-present tense forms:

$$\varepsilon \rightarrow i \;/\; \_\_ \begin{Bmatrix} l \\ m \\ n \\ r \end{Bmatrix} \quad \text{in non-present tense forms}$$

7. Vowel truncation

$$V \rightarrow \emptyset \;/\; \_\_ \left\{ \begin{array}{l} + \left\{ \begin{array}{l} V \\ s \end{array} \right. \\ \# \end{array} \right.$$

Vowel truncation is quite well motivated, although the details of the rule depend on assumptions about the underlying representations more than in the case of most rules. The need for a truncation rule, however, is shown by many alternations: for example, augu 'I grow' vs. audz 'you grow', from /aug + i/.

8. Syncope
The syncope rule converts a sequence of two identical vowels to a long vowel.

$$V + V \rightarrow \bar{V}$$

Both the n/i rule and the i/j rule indicate that it is advantageous to treat surface long vowels as a sequence of identical short vowels. But this treatment requires the syncope rule to convert the vowel sequence to a long vowel.

9. Vowel lengthening
Under rather complicated conditions, the stem of verbs is lengthened.

$$V \rightarrow \bar{V} \ / \ \_\_ \begin{Bmatrix} r \\ l \\ m \\ n \\ u \\ i \end{Bmatrix}$$

in the past-tense forms of verbs with a palatal present-tense infix

celu 'I lift', cēlu 'I lifted'
kaủju 'I kill', kāvu 'I killed'

The remainder of the rules are termed 'lower level' phonological rules by Halle and Zeps, but they do not specify the criteria for this distinction.

10. Spirantization

$$\begin{Bmatrix} t \\ d \end{Bmatrix} \rightarrow \begin{Bmatrix} s \\ z \end{Bmatrix} \ / \ \_\_ \begin{Bmatrix} t \\ d \\ j \end{Bmatrix}$$

metu 'I threw', mest 'to throw', from /met + t/

11. Dental mutation

$$\begin{Bmatrix} c,\ dz \\ s,\ z \\ n,\ l,\ r \end{Bmatrix} \rightarrow \begin{Bmatrix} \check{c},\ d\check{z} \\ \check{s},\ \check{z} \\ \text{ņ},\ \text{ļ},\ \text{ŗ} \end{Bmatrix} \ / \ \_\_ \ j$$

lācis 'bear' nom. sing., lāča 'bear' gen. sing., from /lāc + ja/
(cf. gulbis 'swan' nom. sing., gulbja gen. sing.)

12. j loss

$$j \rightarrow \emptyset \ / \ \left\{ \begin{array}{l} \_\_\ \# \\ \text{palatal consonant } \_\_ \end{array} \right.$$

/lāčj + a/ → lāča

13. Voicing assimilation

All obstruent clusters are either voiceless or voiced, depending on the voicing of the last element.

$$[+\text{obstruent}] \rightarrow [\alpha\text{voice}] \ / \ \_\_ \begin{bmatrix} +\text{obstruent} \\ \alpha\text{voice} \end{bmatrix}$$

The following are some sample derivations of Latvian verbs. In the underlying representation, the verb is composed of a verb stem, an optional tense marker, and a person ending. Many verbs have special tense infixes as well. For example, the -*tt*- in *klīst* is the underlying representation of the traditional -*st*- present-tense infix of Baltic verbs.

lænk + au	
læik + au	n/i rule
liæk + ua	metathesis
liæku	vowel truncation
lieku 'I put'	in the orthography
lænk + æ + i	
lænc + æ + i	k/c rule
læic + æ + i	n/i rule
liæc + i + æ	metathesis
liæc	vowel truncation (morpheme boundaries are inserted to enable the rule to apply twice)
liec 'you put'	in the orthography
kliid + tt + a	
kliid + tt	vowel truncation
klīd + tt	syncope
klīz + st	spirantization
klīst	voicing assimilation (and contraction of identical spirants)
klīst 'he strays'	

Lithuanian Counterparts of Latvian Rules

Before discussing the Lithuanian counterparts of the Latvian rules, it is necessary to say a few words about the underlying representations that have been selected for Lithuanian. In general, the representations of verb stems will be selected to be as close as possible to the Latvian representations, whenever a particular verb has a cognate in Latvian. Long vowels will be analyzed as a sequence of two short vowels, even though this analysis may complicate accent assignment; Lithuanian accent rules will be ignored.

The present tense person endings have been selected on the basis of the person endings that appear with the reflexive verbs, where the endings are protected by a consonant from vowel truncation. There are two sets of past tense person endings. Though these endings are apparently predictable, at least in part, in this paper verbs will simply be considered to be marked [+ -aa past] and [+ -εε past] and be assigned the appropriate person endings on this basis. As in Latvian, many verbs have special tense infixes.

Of the Latvian rules discussed, at least seven also appear in Lithuanian.

1. i/j rule

The i/j rule is identical in Lithuanian and Latvian. For example, the rule is needed in the derivation of the verb *žū́ti* 'perish', with the present tense *žū́sta* and the past tense *žùvo*. The stem can be represented as /žuu-/; the infinitive is formed from /žuu+ ti/. The present tense forms take the -*st*- infix:/žuu + tt + a/→ *žū́sta*. In the past tense forms, the second vowel of the stem precedes another vowel, so the i/j rule applies: /žuu + aa/→ *žùvaa*, and *žûvo* by a later rule. Similarly, *gýti* 'heal' has the present tense formed with a palatal infix:/gii + i + a/→ *gỹja*; in the past tense, the second stem vowel directly precedes another vowel, so the i/j rule applies: /gii + aa/→ *gìjo*.

As in Latvian, *v* and *j* can be regarded as realizations of underlying *u* and *i*; for example, verbs like *dvė̃sti* (*dvė̃sia*, *dvė̃sę*) 'die' can be entered in the lexicon as /dues + ti/; the i/j rule will produce the correct output.

The Lithuanian rule can be formulated to be exactly like the Latvian rule:

$$\begin{Bmatrix} i \\ u \end{Bmatrix} \rightarrow \begin{Bmatrix} j \\ v \end{Bmatrix} \quad / \ \_\_ \ V$$

There are some exceptions to the i/j rule. First, there is the general constraint, shared by Latvian, that the first vowel in a sequence of identical vowels is exempt from the i/j rule. Secondly, a few verbs behave anomalously with respect to the rule; for example, *gúiti* 'chase' keeps both vowels in the infinitive, instead of having the form predicted by the i/j rule: **gviti*. However, since the exceptions appear to be few, they can simply be marked [-i/j rule].

Palatalized and non-palatalized (hard) consonants can contrast only before back vowels; otherwise, consonants are always palatalized before front vowels and hard otherwise. In the traditional orthography, palatalization before back vowels is represented by -*i*-; this device can be employed in the underlying representations as well. For example [k'áušas] 'skull' would have an underlying representation something like /kiauš + as/. The i/j rule would produce /kjauš + as/; consonants preceding *j* or front vowels become palatalized, and the *j* can be dropped. Thus, there is no difficulty with -*i*- as a marker of palatalization. This, of course, simplifies the description of the language, since palatalization can be predicted entirely by rule.

2. i/n rule

The i/n rule has no direct counterpart in Lithuanian, but there are alternations of long vowels with vowel-nasal sequences. For example, *žį́sti*, *žìnda*, *žìndo* 'suck' and *brę́sti*, *brę́sta*, *bréndo* 'mature'. Under rather complex conditions, the nasal of the underlying vowel-nasal sequence vocalizes, creating a sequence of nasalized vowels. Subsequently, all vowels become de-nasalized. Heeschen discusses these alternations, giving the required rule in a form essentially similar to the following:

$$V_1 \, (V_1) \, n \rightarrow V_1 \, (V_1) \, V_1 \; / \; \_\_ \left\{ \begin{array}{l} s, \check{s}, z \; \check{z} \\ e, m, r \end{array} \right.$$

3. Metathesis

The metathesis rule follows the i/j rule, and is also required in Lithuanian phonology. In Latvian, the metathesis rule applies to a great many verbs; in Lithuanian, however, metathesis is a rather minor rule. It can be motivated only for au and ei, not, as far as I can tell, for any of the other sequences which are also subject to metathesis in Latvian. The verb dúoti 'to give' requires both metathesis and the i/j rule in its derivation: /dau + ti/ becomes dúoti and /dau + d + a/ becomes dúoda by metathesis; /dau + εε/ becomes dãvė by the i/j rule.

Some verb stems ending in obstruents have to be entered in the pre-metathesis form to prevent the i/j rule from applying; for example, liẽpti 'to order' would have the underlying representation /leip-/. Thus, the environment for the i/j rule would not be supplied, and metathesis would provide the correct form.

A very large number of verbs, however, are exceptions to metathesis, e.g. kláusti 'ask', geĩsti 'desire', kéikti 'curse', kráuti 'heap up', léisti 'let', etc. Therefore, it may be more economical to mark verb stems to undergo metathesis and to consider the exceptions as normal, rather than to specify the exceptions to metathesis. The metathesis rule would still apply to person endings, however. The unmarked state would be for metathesis to apply to person endings and not to apply to verb stems.

4. Ablaut and Vowel lengthening

Since ablaut and vowel lengthening are both morphologically conditioned rules, the two rules will be discussed together. Lithuanian has an ablaut rule very similar to the Latvian rule:

$$\varepsilon \rightarrow i \; / \; \_\_ \left\{ \begin{array}{l} l \\ m \\ n \\ r \end{array} \right. \quad \text{in non-present tense forms}$$

For example, pir̃kti, per̃ka, pir̃ko 'buy'.

There are at least two rules lengthening vowels. The rule found in Latvian, lengthening vowels in the past tense, also operates in Lithuanian, as exemplified by verbs like: mìnti, mýne 'tread'; pìnti, pýne 'wreathe'; dùrti, dū́re 'stab'; grùmti, grũme 'combat'.

When the stem vowel -a- is lengthened in the past tense, it is subsequently raised to -o-, and, similarly, when -ε- is lengthened, it is raised to -ė-. For example, kárti, kóre 'hang'; pláuti, plóve 'wash'; kélti, kéle 'lift'.

The vowel lengthening rule can be formulated to be very much like the Latvian rule:

$$V \rightarrow \bar{V} \;/\; \_\_ \left\{\begin{matrix} r,\ l \\ m,\ n \\ j,\ v \end{matrix}\right. \quad \text{when the verb takes [+-εε past] tense}$$

All of the verbs showing lengthening in the past tense take the [-εε] past tense, with the exception of eĩti (eĩna, ẽjo) 'go'. It would not be surprising, however, if this verb were irregular, specially marked to undergo the lengthening rule. The rule must also be prevented from applying to verbs like aũti 'to put on shoes' with the past tense form ãvė instead of *ovė, as predicted by the lengthening rule.

Many verbs which show lengthening in the past tense forms also have a nasal present-tense infix, rather than the palatal infix which appears in Latvian, e.g. griáuti, griáuna, grióvė 'thunder'; ráuti, ráuna, róvė, 'tear out'; šáuti, šáuna, šóvė 'shoot'. But this is not true of all verbs showing lengthening in the past tense.

Vowel lengthening takes place in the present tense, rather than in the past, in another set of verbs. All these verbs have -i- or -u- as the stem vowel, and all take the [-aa] past tense endings. For example, dìlti, dýla, dìlo 'wear away'; dùsti, dũsta, dùso 'suffocate'. Apparently, present tense lengthening does not take place before resonants: krìñta 'he falls', mìršta 'he dies'.

The rule can be formulated as follows:

$$\begin{array}{l} V \\ [+\text{high}] \end{array} \rightarrow [+\text{long}] \;/\; \_\_ + \text{Obstruent}$$

in the present tense, when the verb is marked

[+ -aa] past tense

Finally, there is a class of verbs with long stem vowels that lower the stem vowel in the present tense: dė́ti, dẽda, dė́jo 'put'; dvė̃sti, dvẽsia, dvė̃sė 'die'. I can not formulate the rule for vowel lowering, however, because I can not specify the conditions under which the change takes place; some verb stems of essentially identical phonological shape and morphological composition to those listed above do not undergo the rule, e.g. grė́bti, grė́bia, grė́bė 'rake'.

5. Vowel truncation

The vowel truncation rule is difficult to evaluate because, more than other rules, its formulation depends on other components of the analysis. However, the most economical description seems to call for vowel truncation in Lithuanian. In Latvian, of course, vowel truncation is very wide-spread; in fact, loss of vowels in final syllables is one of the major traditionally-cited Latvian innovations.

Vowel truncation in Lithuanian can be motivated if the person endings that show up in the reflexive, where they are protected by a consonant, are considered to appear in the active as well. For example,

lenki + au	
lenkj + au	i/j rule
lenkj + ua	metathesis
lenkj + u	vowel truncation

Finally, the result is lenkiù [leŋk'u] 'I bend'

In the reflexive paradigm, the person endings are protected, and the reflexive form shows the full person ending: lenkiúosi 'I bow (I bend myself)'.

Heeschen formulates the rule quite simply:

$$V \rightarrow \emptyset \ / \ \_\_ \ (s)\#$$

However, he has to exclude the rule from several morphological environments, including the reflexive marker -si, and to postulate extra vowels to protect some endings. Therefore, Lithuanian vowel truncation is not nearly as simple as the rule implies.

Three of the 'lower level' Latvian phonological rules are shared by Lithuanian: spirantization, voicing assimilation, and dental mutation.

6. Spirantization

Lithuanian has a spirantization rule which is identical to the Latvian. For example, /met + ti/ results in mèsti 'to throw'.

7. Voicing assimilation

Similarly, Latvian and Lithuanian share a voicing assimilation rule, assimilating all obstruents in a cluster to the voicing of the last member of the cluster. For example, bégti 'to run' is phonetically [bæ:kti].

8. Dental mutation

The Latvian dental mutation rule has a very limited counterpart in Lithuanian:

$$\left\{\begin{matrix} t \\ d \end{matrix}\right\} \rightarrow \left\{\begin{matrix} \check{c} \\ d\check{z} \end{matrix}\right\} \ \_\_ \ j \begin{bmatrix} V \\ +\text{back} \end{bmatrix}$$

For example skaičiaũ 'I read' and skaiteĩ 'you read'

Lithuanian 'Lower Level' Rules

The verb system of Lithuanian requires a number of 'low level' phonological processes that do not operate in Latvian.

1. Obstruent metathesis

There is an obstruent metathesis rule, exemplified by verbs like blõkšti, blãškia, blõškė 'hit'; and drė̃ksti, drẽskia, drẽskě 'scratch'. Apparently, stem-final spirants and velar stops metathesize. That this metathesis takes place only before consonants is indicated by the following two verbs: blõkšti 'to hit' and bly̌kšti 'to turn pale'. The present tense form of blõkšti is blãškia; it is derived from /blaški + a/ without undergoing obstruent metathesis. The present tense of bly̌kšti, however, is bly̌kšta; the underlying representation is /bliišk + tt + a/. Because the obstruent cluster precedes the -st- infix, the cluster is subject to metathesis. In the past tense, the cluster appears before a vowel, and so appears in the pre-metathesis form: blyško. The rule may be formulated as follows:

velar stop + spirant → spirant + stop / __ + C

2. Nasal metathesis

Seemingly related to obstruent metathesis is metathesis of the nasal 'infix' with the last element of the stem when the stem ends in an obstruent. This is exemplified by verbs like the following: krìsti, kriñta, krìto 'fall'; (pa-) tìkti, tiñka, tìko, 'like'; klùpti, klum̃pa, klùpo 'trip'. The simplest way to handle this phenomenon is to assume that the nasal infix is added to the stem, metathesizes when it follows an obstruent, and then assimilates to the position of articulation of the following obstruent. A sample derivation would be the following:

klup + N + a	
kluNpa	nasal metathesis
klumpa	assimilation
klum̃pa 'he trips'	

If the nasal infix is not followed by an obstruent, i.e. in a present tense form like pláuna 'he washes', the nasal infix is realized as -n-.

The following rules are required:

Obst. + N → N + Obst. / ___ +

N → n / __ t, d
     m / __ p, b
     ŋ / __ g, k

N → n

3. Vowel raising

As mentioned before, non-nasal -aa- becomes long o and -εε- becomes long ė. The syncope rule, which is also required in Lithuanian, and vowel de-nasalization, both 'clean-up' rules, would be ordered after vowel raising.

4. Palatalization

There is very wide-spread palatalization of consonants in Lithuanian; any consonant becomes palatalized in the appropriate environment, even non-native consonants in borrowed words. For example, filològas 'philologist' and fìzika 'physics' both have palatalized f. The rule for palatalization is:

$$C \rightarrow C' \ / \ \underline{\quad} \ \left\{\begin{array}{l}\text{front V}\\ j\\ C'\end{array}\right.$$

5. Spirant assimilation

Dental spirants become palatal spirants before palatal affricates:

$$\left\{\begin{matrix} s \\ z \end{matrix}\right\} \rightarrow \left\{\begin{matrix} \check{s} \\ \check{z} \end{matrix}\right\} \quad / \quad \_\_ \left\{\begin{matrix} \check{c} \\ d\check{z} \end{matrix}\right\}$$

This is clearly indicated by a form like *pė́sčias* 'on foot' which is phonetically [peš'č'as].

6. Final devoicing

Consonants are devoiced and de-palatalized in word-final position:

$$C \rightarrow \begin{bmatrix} -\text{voice} \\ -\text{sharp} \end{bmatrix} \quad / \quad \_\_\#$$

Conclusion

As is clear from the discussion of individual rules, there are three possible relations between the rules in the two languages: the rules are identical in the two languages, a rule has no counterpart in the other language, or a rule has changed in some way.

Four rules appear to be identical in the two languages: the i/j rule, vowel lengthening in the past tense, spirantization, and voicing assimilation. The i/j rule and vowel lengthening are best considered to be inherited rules, operating at a high level in the phonology. Spirantization and voicing assimilation, however, are both low level phonological rules; voicing assimilation is preceded by several other innovative low level rules in Lithuanian, e.g. final devoicing precedes voicing assimilation.

It is tempting to speculate that the status of the two sets of rules is not the same. Though the claim can not be substantiated here, it may be that a certain set of rules should be viewed as defining constraints on the shape of the phonological output, rather than defining phonological alternations. The spirantization and voicing assimilation rules appear to be of this 'lower level' type.

There are five rules that appear in both languages but not in exactly the same form. These are the n/i rule, ablaut, metathesis, vowel truncation, and dental mutation. Only dental mutation is a 'lower level' rule; the other four rules are higher-level phonological rules. In all cases, the Latvian rules appear to be simpler, in one way or another, than the Lithuanian rules.

Assuming that the Latvian n/i rule is an extension of the Lithuanian rule defining long-vowel, vowel-nasal alternations, the Latvian rule has been simplified in two ways. In Lithuanian, the vocalized nasal must match the preceding vowel in all features; in Latvian, the vocalized nasal is always a high vowel, matching only in the front-back dimension. Secondly, the Latvian rule specifies a simpler environment, before any consonant, rather than the rather complicated set of consonants required for the Lithuanian rule.

As is clear from the preceding discussion of metathesis, the rule not only applies to more sequences of vowels but also to more stems in Latvian than in Lithuanian. Latvian, therefore, has generalized the applicability of the rule.

If ablaut and other morphologically conditioned alternations are considered together, then it is quite clear that the Lithuanian system is more complex. It includes not only the two rules that appear in Latvian but also several others: it involves more different kinds of alternations and more complicated rules to define them.

Vowel truncation is much more restricted in Lithuanian than in Latvian. As has been mentioned previously, virtually all vowels in final syllables have disappeared in Latvian, but this is by no means the case in Lithuanian. Apparently, Latvian has extended the applicability of the rule.

Finally, the dental mutation rule, assuming that it is basically the same rule in the two languages, applies to almost the whole class in Latvian but to only two members of the class in Lithuanian.

Some rules appear in only the one or the other languages. The various morphologically-conditioned lengthening rules of Lithuanian have already been mentioned; these rules are historical retentions in Lithuanian which are lost in Latvian. The status of the two Lithuanian consonant metathesis rules is not clear; with the data presently at my disposal, I could not determine whether the rules are innovations or retentions in Lithuanian. Palatalization and final devoicing are both clearly Lithuanian innovations, probably additions to the set of 'output condition' rules.

Latvian seems to have innovated two rules: the k/c rule and the ε/e rule. These innovations are problematic, however, in that both these rules appear at a rather early stage of the phonology. The k/c rule and the ε/e rule must precede both vowel truncation and metathesis. For example, the environment required for the k/c rule may be deleted by vowel truncation: audz 'you grow', from /aug + ε + i/, vs. aug 'he grows', from /aug + a/. Secondly, a form like vilki 'wolves', from /vilk + ai/, indicates that the k/c rule precedes metathesis, since the k/c rule is inapplicable when k precedes a front vowel because of metathesis. That the ε/e rule precedes vowel truncation is clear in the derivation of mest [mest] 'to throw', from /mεt + ti/; forms that do not have a high vowel in the inflectional suffixes keep ε: metu [mεtu] 'I throw' and met [mεt] 'he throws'.

It is not clear exactly how the two rules came to be ordered early in the grammar. Recently there has been considerable discussion about rule insertion, summarized in King (1970). King concludes that rule insertion--the addition of a rule which must be ordered before a phonological rule present in an earlier stage of the grammar--is a possible type of linguistic change, but that there are very few good examples of it. At first glance, the Latvian k/c and ε/e rules look like examples of rule insertion; however, the rules may also appear in their present order because of rule reordering. The two rules are crucially ordered only with respect to vowel truncation and metathesis, both rules that have been greatly generalized in Latvian. It is possible that the k/c and ɛ/e rules appear early in the grammar because of reordering from 'bleeding' to 'feeding' order. In an earlier stage of Latvian, the rules would apply in the following order: metathesis, vowel truncation, k/c rule and ɛ/e rule. As vowel truncation became generalized, more and more environments for the k/c rule and ɛ/e rule were eliminated by the deletion of final vowels; the rules now operated in 'bleeding' order. At this point, the rules were reordered to 'feeding' order. To determine which process, rule insertion or reordering, is responsible for the present rule order in Latvian, more evidence is necessary than is available to me at the moment.

The relationship of the rules in the two languages can be summarized in the following way. Latvian has simplified rules, generalized their application, and added two high-level rules; Lithuanian has retained complex rules which apply under complicated circumstances, and added low-level rules.

The judgment that Latvian is innovating and Lithuanian conservative is interesting in this context. Lithuanian preserves complex alternations but rather freely changes their phonetic realization; Latvian changes the phonetic realization much less, but loses complex alternations. The observation is slightly trivial but still worth making: a conservative vs. an innovating phonology is not defined in terms of surface phonetic realization.

Obviously, the rules discussed in this paper represent only a small fragment of Lithuanian and Latvian phonology. It seems, however, that a comparison of the phonological systems of the two languages can provide very interesting material for a study of language change.

Bibliography

Halle, M. and V. Zeps. "Survey of Latvian Morphophonemics", Quarterly Progress Report 83, Research Laboratory in Electronics, M.I.T., Cambridge, Massachusetts. 1966.

Heeschen, C. "Lithuanian Morphophonemics", Quarterly Progress Report 85, Research Laboratory in Electronics, M.I.T., Cambridge, Massachusetts. 1967.

Dambriūnas, Leonardas, Antanas Klimas and William R. Schmalstieg. Introduction to Modern Lithuanian. Brooklyn, New York: Franciscan Fathers Press. 1966.

King, Robert D. "Can Rules be Added in the Middle of Grammars?", Forum Lecture, Ohio State University Linguistic Institute, August 4, 1970.

Pilka, D. Lietuvių Kalbos Gramatika. South Boston, Mass. 1939.

Senn, Alfred. Handbuch der litauischen Sprache. Heidelberg: Carl Winter. 1966.

Zeps, Valdis J. "Base Shapes of Latvian Morphemes", in Baltic Linguistics, Thomas F. Magner and William R. Schmalstieg, eds., University Park, Penn.: Pennsylvania State Univ. Press. 1970.